



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 123041

TO: Georgia L Helmer
Location: REM-2C15&2C18
Art Unit: 1638
Wednesday, June 02, 2004

Case Serial Number: 10/015637

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-B55
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Helmer,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
CM-1, Rm. 6-A-06
605-1155

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:15:29 ; Search time 4648 Seconds

(without alignments)
10705.207 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148
Sequence: 1 taggaccccttcacatagaanaa.....agagagtcgacgctaatcga 1148Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rod:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	1148	AX463281	AX463281 Sequence
2	1148	100.0	1832	AX463294	AX463294 Sequence
3	1144.8	99.7	1821	AX439309	AX439309 Sequence
4	1144.8	99.7	1390	PVARSX1	Z50202 P. vulgaris
5	1122	97.7	1122	AX463282	AX463282 Sequence
6	795.6	69.4	1866	AX463293	AX463293 Sequence
7	795	69.3	1872	AX463292	AX463292 Sequence
8	274.6	23.9	4564	PHVARC1A	ME8913 Phaseolus v
9	139.4	12.1	2288	AFI93029	AFI93029 Phaseolus
10	118.2	10.3	1768	PVPDLECI	X04609 P. vulgaris
11	109.2	9.5	22243	PPAVAR23A	I40609 Plasmodium
12	108.4	9.4	1395	PVDLECI	X02408 Phaseolus v
13	108.4	9.4	1441	PHVDRBCA	K03288 P. vulgaris
14	95.8	8.3	24631	AC111404	AC111404 Rattus no
15	95.2	8.3	93791	AC138073	AC138073 Homo sapi
16	95	8.3	125933	AC146394	AC146394 Pan trogl
17	92	8.0	1689	PHYLECT	J01261 P. vulgaris
18	92	8.0	4846	AF325188	AF325188 Phaseolus
19	91.8	8.0	111861	AC069435	AC069435 Homo sapi
20	91.8	8.0	170627	AC125567	AC125567 Rattus no
21	91.6	8.0	13932	PVPDLECI	X04659 P. vulgaris
22	91.2	7.9	175544	AC117342	AC117342 Rattus no
23	90.6	7.9	258658	AE014832	AE014832 Plasmodium
24	90	7.8	8056	AX599046	AX599046 Sequence
25	90	7.8	110000	PFMAL8P1_12	Conjunction (13 o
26	89.8	7.8	810	AE001398	AE001398 Arabidops
27	89.8	7.8	14867	AE001398	AE001398 Arabidops
28	89.8	7.8	136688	AL353783	AL353783 Human DNA
29	89.6	7.8	103344	HS1100E15	AL355551 Human DNA
30	89.6	7.8	141275	BX510640	BX510640 Dario rer
31	89.6	7.8	171333	AC127704	BX248097 Zebrafish
32	89.6	7.8	249867	AC127704	AC127704 Rattus no
33	88.8	7.7	108908	PFMAL3P8	AL034560 Plasmodium
34	88.8	7.7	195620	BX088600	BX088600 Dario rer
35	88.8	7.7	250029	AE014820	AE014820 Plasmodium
36	88.6	7.7	250743	AE014836	AE014836 Plasmodium
37	87.4	7.6	1434	AE014830	AE014830 Arabidops
38	87.4	7.6	2426	S0U49822	U49822 Saccharomyc
39	87.4	7.6	170880	BX640469	BX640469 Dario rer
40	87.2	7.6	250029	AE014830	AE014830 Plasmodium
41	87.2	7.6	1453	AJ591978	AJ591978 Arabidops
42	87.2	7.6	164640	BX005461	BX005461 Dario rer
43	87.2	7.6	168558	BX322549	BX322549 Dario rer
44	87	7.6	3683	AX598999	AX598999 Sequence
45	87	7.6	136657	AC068600	AC068600 Homo sapi

ALIGNMENTS

RESULT 1	AX463281	1148 bp	DNA	linear	PAT 15-JUL-2002
LOCUS	AX463281	Sequence 1 from Patent WO0250295.			
DEFINITION	AX463281				
ACCESSION	AX463281				
VERSION	AX463281.1	GI:21886632			
SOURCE	Phaseolus vulgaris				
ORGANISM	Phaseolus vulgaris				
REFERENCE	1	Ulmassov, T., Wang, Q., Dubois, P. and Liang, J.			
AUTHORS					

TITLE Arcelin-5 promoter and uses thereof
JOURNAL Patent: WO 0250295-A 1 27-JUN-2002;
RENESEN LLC (US)
FEATURES Location/Qualifiers
source 1..1148
/organism="Phaseolus vulgaris"
/mol_type="unassigned DNA"
/db_xref="taxon:3885"

ORIGIN

Query Match 100.0%; Score 1148; DB 6; Length 1148;
Best Local Similarity 100.0%; Pred. No. 8.4e-173;
Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGATCTTCATAGAAAATGCTTATTTCTCCATACACAGCAAGGGCAACGTTA 60
DB 1 TAGGATCTTCATAGAAAATGCTTATTTCTCCATACACAGCAAGGGCAACGTTA 60
QY 61 ACAAAACAATTATGTTTCATTTGAGATTAGAGAGTAAGAGAAAAAGATTAAA 120
DB 61 ACAAAACAATTATGTTTCATTTGAGATTAGAGAGTAAGAGAAAAAGATTAAA 120
QY 121 AAAATGCTTATCTCTTTGTTTCTGTAATATATATATAGACTTAACTTTATAT 180
DB 121 AAAATGCTTATCTCTTTGTTTCTGTAATATATATATAGACTTAACTTTATAT 180
QY 181 AATATATGTAATAGTTTCTAGTCAAGACACACTCAGACAGAGATTCAAGAAA 240
DB 181 AATATATGTAATAGTTTCTAGTCAAGACACACTCAGACAGAGATTCAAGAAA 240
QY 241 CAATTTGTTAAACATCTTATAGAACTTTAGTTAGTCTTGAAGTTAGATTAAACA 300
DB 241 CAATTTGTTAAACATCTTATAGAACTTTAGTTAGTCTTGAAGTTAGATTAAACA 300
QY 301 AAAAAATTACACAGAGAAACAATAAACCACTACCGTCAAGTTATCATAGATGA 360
DB 301 AAAAAATTACACAGAGAAACAATAAACCACTACCGTCAAGTTATCATAGATGA 360
QY 361 AATGTTTGAATGATTAATATATATACACACAAAAATACATCTTATATACAAATAT 420
DB 361 AATGTTTGAATGATTAATATATATATACACACAAAAATACATCTTATATACAAATAT 420
QY 421 GTTATACATATATTTTGTAAAACTTGAAGTTTCTCAAACTTTCAATATACAGATTAG 480
DB 421 GTTATACATATATTTTGTAAAACTTGAAGTTTCTCAAACTTTCAATATACAGATTAG 480
QY 481 AGTTATGAAATCAAAATATTTAAAAAATATATTTAAAAAACAATTCTAAAGTCATT 540
DB 481 AGTTATGAAATCAAAATATTTAAAAAATATATTTAAAAAACAATTCTAAAGTCATT 540
QY 541 CAGATCTCTCAACACTGTGTGATCATTTAGTCAATGATGATGATCAATCTTGTAGTTC 600
DB 541 CAGATCTCTCAACACTGTGTGATCATTTAGTCAATGATGATGATCAATCTTGTAGTTC 600
QY 601 ACAACAGGTAATAATTAAGATTAAGTAACTAGGGGAATATATATATATATCAATTAAAT 660
DB 601 ACAACAGGTAATAATTAAGATTAAGTAACTAGGGGAATATATATATATATCAATTAAAT 660
QY 661 AAAAAAGGAAAAATCAAAATTGAAATTTTGAATCCCAATGACACAACTCACATGAC 720
DB 661 AAAAAAGGAAAAATCAAAATTGAAATTTTGAATCCCAATGACACAACTCACATGAC 720
QY 721 GCTGCACTCTGACCTCCCTCTCTCCACACAGTCTCATCTGACCTTGGAGCTT 780
DB 721 GCTGCACTCTGACCTCCCTCTCTCCACACAGTCTCATCTGACCTTGGAGCTT 780
QY 781 TTCACTATGACACAACTGCGCATGATGTGCAAGTGAAGCTCTTCTCTCCCATGAT 840
DB 781 TTCACTATGACACAACTGCGCATGATGTGCAAGTGAAGCTCTTCTCTCCCATGAT 840
QY 841 GACACCACTGGGCGATGATGCTGCGCACTGAGCTTCCCACTCTCTCTCATATAGCCCTAC 900
DB 841 GACACCACTGGGCGATGATGCTGCGCACTGAGCTTCCCACTCTCTCTCATATAGCCCTAC 900

QY 901 TGGCCATGACACATGCGACCTGACACATCTCTCACTTCCCATTTGCTACCTGCAACCG 960
DB 901 TGGCCATGACACATGCGACCTGACACATCTCTCACTTCCCATTTGCTACCTGCAACCG 960
QY 961 CTCTCTCCATTAATATCTATTTAAATTTAAACAAATATTTGATATCTTTTGTATGA 1020
DB 961 CTCTCTCCATTAATATCTATTTAAATTTAAACAAATATTTGATATCTTTTGTATGA 1020
QY 1021 CGTGAATGATTCGATGCTGTTTATATATATTTGATATTTGAGTGAATATATAATGA 1080
DB 1021 CGTGAATGATTCGATGCTGTTTATATATATTTGATATTTGAGTGAATATATAATGA 1080
QY 1081 AAGAAAAAGTTGAAAGATTTTGTGATTTGTTGATATATATAGAGAAAGAGTGATG 1140
DB 1081 AAGAAAAAGTTGAAAGATTTTGTGATTTGTTGATATATATAGAGAAAGAGTGATG 1140
QY 1141 TTAATGCA 1148
DB 1141 TTAATGCA 1148

RESULT 2
AX463294 1832 bp DNA linear PAT 15-JUL-2002
LOCUS
DEFINITION
Sequence 14 from Patent WO0250295.
AX463294
VERSION
AX463294.1 GI:21886245
KEYWORDS
SOURCE
ORGANISM
Phaseolus vulgaris
Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE
1 Oulassov,T., Wang,Q., Dubois,P. and Liang,J.
ATTHORS
Arcelin-5 promoter and uses thereof
JOURNAL
Patent: WO 0250295-A 14 27-JUN-2002;
RENESEN LLC (US)
FEATURES Location/Qualifiers
source 1..1832
/organism="Phaseolus vulgaris"
/mol_type="unassigned DNA"
/db_xref="taxon:3885"

ORIGIN

Query Match 100.0%; Score 1148; DB 6; Length 1832;
Best Local Similarity 100.0%; Pred. No. 7.7e-173;
Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGATCTTCATAGAAAATGCTTATTTCTCCATACACAGCAAGGGCAACGTTA 60
DB 659 TAGGATCTTCATAGAAAATGCTTATTTCTCCATACACAGCAAGGGCAACGTTA 718
QY 61 ACAAAACAATTATGTTTCATTTGAGATTAGAGAGTAAGAGAAAAAGATTAAA 120
DB 719 ACAAAACAATTATGTTTCATTTGAGATTAGAGAGTAAGAGAAAAAGATTAAA 120
QY 121 AAAATGCTTATCTCTTTGTTTCTGTAATATATATATAGACTTAACTTTATAT 180
DB 779 AAAATGCTTATCTCTTTGTTTCTGTAATATATATATAGACTTAACTTTATAT 180
QY 181 AATATATGTAATAGTTTCTAGTCAAGACACACTCAGACAGAGATTCAAGAAA 240
DB 839 AATATATGTAATAGTTTCTAGTCAAGACACACTCAGACAGAGATTCAAGAAA 240
QY 241 CAATTTGTTAAACATCTTATAGAACTTTAGTTAGTCTTGAAGTTAGATTAAACA 300
DB 899 CAATTTGTTAAACATCTTATAGAACTTTAGTTAGTCTTGAAGTTAGATTAAACA 300
QY 301 AAAAAATTACACAGAGAAACAATAAACCACTACCGTCAAGTTATCATAGATGA 360
DB 301 AAAAAATTACACAGAGAAACAATAAACCACTACCGTCAAGTTATCATAGATGA 360

Db 959 AAAAAATTCACACGAGAAACAACTAACCCCTACCGTACGTTATCATAGATCA 1018
 Qy 361 AATGTTTGAATATCATTTAAATTAACACACACAAAAATACATCTAATTAATTAACAATATAT 420
 Db 1039 AATGTTTGAATATCATTTAAATTAACACACAAAAATACATCTAATTAATTAACAATATAT 1078
 Qy 421 GTTATACATATATTTTGTAAAAAATTAAGTTTTCAAAATTAATTAATTAATTAATTAAG 480
 Db 1079 GTTATACATATATTTTGTAAAAAATTAAGTTTTCAAAATTAATTAATTAATTAATTAAG 1138
 Qy 481 AGTTATATAGAAATACAAATATTTTAAATATATATTTTAAATTAATTAATTAATTAAG 540
 Db 1139 AGTTATATAGAAATACAAATATTTTAAATATATATTTTAAATTAATTAATTAATTAAG 1198
 Qy 541 CAGATCTCTCAACCTGTGTGATCTTTTGTATATATATATATATATATATATATATATAT 600
 Db 1199 CAGATCTCTCAACCTGTGTGATCTTTTGTATATATATATATATATATATATATATATAT 1258
 Qy 601 ACAACAGATTAATTAATTAAGATTAACCTAGGAAATATATATATATATATATATATATAT 660
 Db 1259 ACAACAGATTAATTAATTAAGATTAACCTAGGAAATATATATATATATATATATATATAT 1318
 Qy 661 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 720
 Db 1319 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1378
 Qy 721 GGTGCACCTCAACCTGTGTGATCTTTTGTATATATATATATATATATATATATATATAT 780
 Db 1379 GGTGCACCTCAACCTGTGTGATCTTTTGTATATATATATATATATATATATATATATAT 1438
 Qy 781 TTCACTATGACACAACTGCCATGATGTCACCTGAGCTCTTCTCTTCCATGAT 840
 Db 1439 TTCACTATGACACAACTGCCATGATGTCACCTGAGCTCTTCTCTTCCATGAT 1498
 Qy 841 GACACCACTGGGATGATGATGTCACCTGAGCTCTTCTCTTCCATGATGATG 900
 Db 1499 GACACCACTGGGATGATGATGTCACCTGAGCTCTTCTCTTCCATGATGATG 1558
 Qy 901 TGGCCATGACACCTGCCATGATGTCACCTGAGCTCTTCTCTTCCATGATGATG 960
 Db 1559 TGGCCATGACACCTGCCATGATGTCACCTGAGCTCTTCTCTTCCATGATGATG 1618
 Qy 961 CTTCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
 Db 1619 CTTCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1678
 Qy 1021 CGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 Db 1679 CGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1738
 Qy 1081 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1140
 Db 1739 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1798
 Qy 1141 TTAATGCA 1148
 Db 1799 TTAATGCA 1806

RESULT 3
 AX343909 1821 bp DNA linear PAT 01-FEB-2002
 LOCUS AX343909
 DEFINITION Sequence 1 from Patent WO0200899.
 ACCESSION AX343909
 VERSION AX343909.1 GI:18491955
 KEYWORDS
 SOURCE
 ORGANISM
 Phaseolus vulgaris
 Phaseolus vulgaris
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Phaseolus.
 REFERENCE
 1

AUTHORS Angenon, G., de Jaeger, G., Goossens, A. and Depicker, A.
 TITLE Heterologous gene expression in plants
 JOURNAL Patent: WO 0200899-A1 03-JAN-2002;
 Vilems Interuniversitair Instituut voor Biotechnologie vzw. (BE)
 FEATURES
 source
 1. 1821
 /organism="Phaseolus vulgaris"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3885"

Query Match 99.74; Score 1144.8; DB 6; Length 1821;
 Best Local Similarity 99.8%; Pred. No. 2.5e-172;
 Matches 1146; Conservatives 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGATCTCTCAACCTGTGTGATCTTTTGTATATATATATATATATATATATATATATAT 60
 Db 674 TGGATCTCTCAACCTGTGTGATCTTTTGTATATATATATATATATATATATATATATAT 733
 Qy 61 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 120
 Db 734 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 793
 Qy 121 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 180
 Db 794 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 853
 Qy 181 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 240
 Db 854 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 913
 Qy 241 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 300
 Db 914 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 973
 Qy 301 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 360
 Db 974 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1033
 Qy 361 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 420
 Db 1034 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1093
 Qy 421 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 480
 Db 1094 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1153
 Qy 481 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 540
 Db 1154 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1213
 Qy 541 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 600
 Db 1214 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1273
 Qy 601 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 660
 Db 1274 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1333
 Qy 661 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 720
 Db 1334 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1393
 Qy 721 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 780
 Db 1394 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1453
 Qy 781 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 840
 Db 1454 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 1513
 Qy 841 AAAAAAGGAAATCAATTAAGATTTTGTATTTTCCCAATGACACAACTCAACATGAC 900

Db 1514 GACACCACTGGGCGATGCACTGCGACCTGACCTCCACCTCTTCTCATTTAGAGCTTAC 1573
 Qy 901 TGGCCATGACACATGCGACCTGACCACTCTCTCTACCTCCATTCCTGCGCAACCG 960
 Db 1574 TGGCCATGACACATGCGACCTGACCACTCTCTCTACCTCCATTCCTGCGCAACCG 1633
 Qy 961 CT 1020
 Db 1634 CT 1693
 Qy 1021 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 Db 1694 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1753
 Qy 1081 AAGAAAAAGTTGGAAGATTGCAATTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
 Db 1754 AAGAAAAAGTTGGAAGATTGCAATTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1813
 Qy 1141 TTAATGCA 1148
 Db 1814 TTAATGCA 1821

RESULT 4
 PVARC5X1
 LOCUS P.vulgaris arcs-1 gene. 3900 bp DNA linear PLN 21-AUG-1998
 DEFINITION
 VERSION 250202.1 GI:3451281
 KEYWORDS arcs-1 gene; arcelin 5a.
 SOURCE Phaseolus vulgaris
 ORGANISM Baktaryca; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.

REFERENCE 1
 AUTHORS Goossens, A., Geremia, R., Bauw, G., Van Montagu, M. and Angenon, G.
 TITLE Isolation and characterization of arcelin-5 proteins and cDNAs
 JOURNAL Eur. J. Biochem. 225 (3), 787-795 (1994)
 MEDLINE 9504555
 PUBMED 7957215
 REMARK (sites)

REFERENCE 2
 AUTHORS Goossens, A., Ardiles Diaz, W., De Keyser, A., Van Montagu, M. and Angenon, G.
 TITLE Nucleotide sequence of an arcelin-5 genomic clone from wild
 JOURNAL Phaseolus vulgaris
 REMARK Phaseolus vulgaris
 AUTHORS plant Physiol. 109, 722-722 (1995)
 JOURNAL (sites)

REFERENCE 3
 AUTHORS Goossens, A.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-1995) Goossens A., Universiteit Gent, Laboratorium Genetika, K.L.Ledeganckstraat 35, Gent, Belgium, 9000
 REMARK Revised by (4)

REFERENCE 4
 AUTHORS Goossens, A.
 TITLE Direct Submission
 JOURNAL Submitted (13-AUG-1998) Goossens A., Universiteit Gent, Laboratorium Genetika, K.L.Ledeganckstraat 35, Gent, Belgium, 9000
 COMMENT On Aug 25, 1998 this sequence version replaced gi:299816.
 FEATURES
 SOURCE

1. 3900
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /strain="G02771"
 /db_xref="taxon:3885"
 /clone="GARC5-8A"
 /haplotype="arcs"
 /tissue_type="leaf"
 /1707..1710
 /citation=(2)
 CAAT_signal

TATA_signal 1790..1793
 TATA_signal /citation=(2)
 gene 1792..1797
 gene /citation=(2)
 mRNA 1822..2754
 mRNA /gene="arcs-1"
 /citation=(2)
 /evidence=experimental
 5'UTR 1822..1834
 /gene="arcs-1"
 /citation=(2)
 /evidence=experimental
 gene 1835..2620
 /gene="arcs-1"
 /citation=(2)
 /evidence=experimental
 CDS 1835..2620
 /gene="arcs-1"
 /citation=(2)
 /evidence=experimental
 /product="arcelin 5a"
 /protein_id="CA90585.1"
 /db_xref="GI:3451282"
 /db_xref="GOA:Q42460"
 /translation="MASSKLSLALFLVLTTHANSATETS ENF PNFHTDKLLQNGA
 TTSKGLQINGVSNELPRVDSIGAFYSDPIQIDSNVASFVNTFTPIIRAKQS
 ISAVGLAPVAPVNSPPOKQEFIGIPYNNPDEPARVAVVNTFKRIDPKKFIK
 PYVNECDPHKNGEKTDVQITVDSNDLSVFLAPFTYQVRCVSATVHLKEVDEN
 VSVGFSATSGALTEDTTEHDVLSWFSKFRKUSNLIANNIL"
 /1835..1897
 /gene="arcs-1"
 /citation=(1)
 /evidence=experimental
 mat_peptide 1898..2617
 /gene="arcs-1"
 /product="arcelin 5a"
 /citation=(1)
 /evidence=experimental
 3'UTR 2618..2754
 /gene="arcs-1"
 /citation=(2)
 /evidence=experimental
 polyA_signal 2704..2709
 /gene="arcs-1"
 /citation=(1)
 polyA_signal 2708..2713
 /gene="arcs-1"
 /citation=(1)
 polyA_signal 2722..2727
 /gene="arcs-1"
 /citation=(1)
 polyA_signal 2726..2731
 /gene="arcs-1"

ORIGIN

Query Match 99.7% Score 1144.8; DB 8; Length 3900;
 Best Local Similarity 99.8%; Pred. No. 2,2e-172;
 Matches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 TGGATCTCTCAATGAAAAATGTTATTTCTCTATCCACGACAAAGGGGCAACGTTA 60
 Db 674 TGGATCTCTCAATGAAAAATGTTATTTCTCTATCCACGACAAAGGGGCAACGTTA 733
 Qy 61 ACAAACAATTTAGTTTCATTGAGATTAGAGAGTTAGAGAGAAAAAGATTAAAA 120
 Db 734 ACAAACAATTTAGTTTCATTGAGATTAGAGAGTTAGAGAGAAAAAGATTAAAA 793
 Qy 121 AAAATGCTTATCTCTTGTGTTGTTGTAATTAATTAATTAATTAATTAATTAAT 180
 Db 794 AAAATGCTTATCTCTTGTGTTGTTGTAATTAATTAATTAATTAATTAATTAAT 853
 Qy 181 AATTAATGTAATTAAGTTTTCATGATGAGACCAACCTCAGAGCAAGATTTCAGAAAA 240

```

Db      854 AATAATTGTAATAGGTTTCTAGTCATGACACCACTGAGACAGAGTTTCAAGAAA 913
Qy      241 CAATTTGTTAAACATCTTATTAGAAACCTTTAGTTAACTGTGAAGTTAGATTAAACA 300
Db      914 CAATTTGTTAAACATCTTATTAGAAACCTTTAGTTAACTGTGAAGTTAGATTAAACA 973
Qy      301 AAAAAAATTGACACGAGAAACACATTAACCCACTGACGTTATCATPAGATCA 360
Db      974 AAAAAAAGTACACGAGAAACACATTAACCCACTGACGTTATCATPAGATCA 1033
Qy      361 AATGTTGATATCATTAATTAATTAACACAAAAATCATCTATTTAACTAACTAT 420
Db      1034 AATGTTGATATCATTAATTAATTAACACAAAAATCATCTATTTAACTAACTAT 1093
Qy      421 GTTATACATATATTTTGTAAAAAATTAGAGTTTCAAAACATTTAAATACATGAT 480
Db      1094 GTTATACATATATTTTGTAAAAAATTAGAGTTTCAAAACATTTAAATACATGAT 1153
Qy      481 AGTTTATAGAAATACAAATTTTAAAAATATATTTTAAAAAATCTTAAGCAT 540
Db      1154 AGTTTATAGAAATACAAATTTTAAAAATATATTTTAAAAAATCTTAAGCAT 1213
Qy      541 CAGATCTCTCACACCTGTGTGATCATTTAGTATGTATGTATGATCATTTAGTTC 600
Db      1214 CAGATCTCTCACACCTGTGTGATCATTTAGTATGTATGTATGATCATTTAGTTC 1273
Qy      601 ACAACAGAGTAAATTAATTAAGATTAACCTGAGAAATATATATTAATTAAT 660
Db      1274 ACAACAGAGTAAATTAATTAAGATTAACCTGAGAAATATATATTAATTAAT 1333
Qy      661 AAAAAAGGAAATTAATTAAGATTTTGTATTTCCCATGACGACAACTCAGATGAC 720
Db      1334 AAAAAAGGAAATTAATTAAGATTTTGTATTTCCCATGACGACAACTCAGATGAC 1393
Qy      721 GGTGCACTCTCAGCTCCTCTCTCTCCACATGTCTCATGTCTTCACTTGGCTTT 780
Db      1394 GGTGCACTCTCAGCTCCTCTCTCTCCACATGTCTCATGTCTTCACTTGGCTTT 1453
Qy      781 TTCACTATGACACAACTGCGCATGACATGTGCAAGTGGCTCCTCTCCCATGAT 840
Db      1454 TTCACTATGACACAACTGCGCATGACATGTGCAAGTGGCTCCTCTCCCATGAT 1513
Qy      841 GACACCACTGCGCATGACATGTGCAAGTGGCTCCTCTCTCATTAATGAGCTTAC 900
Db      1514 GACACCACTGCGCATGACATGTGCAAGTGGCTCCTCTCTCATTAATGAGCTTAC 1573
Qy      901 TGGCCATGACACCTGCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db      1574 TGGCCATGACACCTGCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1633
Qy      961 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db      1634 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1693
Qy      1021 CGTGAATGACATTCGATCGTTGTAAATTTGTAATTTGAGTTGAATTAATTA 1080
Db      1694 CGTGAATGACATTCGATCGTTGTAAATTTGTAATTTGAGTTGAATTAATTA 1753
Qy      1081 AAAAAAAGTTGGAAGATTTTGCATTTGTGTGATTAATTAAGAAAGAGTATG 1140
Db      1754 AAAAAAAGTTGGAAGATTTTGCATTTGTGTGATTAATTAAGAAAGAGTATG 1813
Qy      1141 TTAATGCA 1148
Db      1814 TTAATGCA 1821

```

RESULT 5
 AX463282
 LOCUS
 DEFINITION Sequence 2 from Patent WO0250295.
 AX463282
 ACCESSION
 1122 bp DNA linear PAT 15-JUL-2002

```

VERSION AX463282.1 GI:2186233
KEYWORDS
SOURCE
ORGANISM
  Phaseolus vulgaris
  Phaseolus vulgaris
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
  Phaseolus.
REFERENCE
  1 Oulmasov, T., Wang, O., Dubois, P. and Liang, J.
  Arcelin-5 promoter and uses thereof
  Patent: WO 0250295-A 27-JUN-2002;
  REMSENSEN LLC (US)
FEATURES
  source
    location/Qualifiers
    1..1122
    /organism="Phaseolus vulgaris"
    /mol_type="unassigned DNA"
    /db_xref="taxon:3855"
ORIGIN
  Query Match 97.7%; Score 1122; DB 6; Length 1122;
  Best Local Similarity 100.0%; Pred. No. 1,1e-168; Mismatches 0; Gaps 0;
  Matches 1122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  3 GATTCCTCAATGAGAAATGTTGTTATTTCTCATCAACACAAAGGGGCAACAGTTAC 62
  1 GATCCTTCATAGAAATGTTGTTATTTCTCATCAACACAAAGGGGCAACAGTTAC 60
  63 AAAAAAATTTATGTTGTTATTTGTTATTTGTTATTTGTTATTTGTTATTTGTTAT 122
  61 AAAAAAATTTATGTTGTTATTTGTTATTTGTTATTTGTTATTTGTTATTTGTTAT 120
  123 AATGCTTATCTCTTGTGTTCTGTAATTAATTAATTAAGACTTAACTTTAATATA 182
  121 AATGCTTATCTCTTGTGTTCTGTAATTAATTAATTAAGACTTAACTTTAATATA 180
  183 TAATTTATTTGTTGTTTCTGATCAAGACACATCTGAGACAGAGTTTCAAGAAACA 242
  181 TAATTTATTTGTTGTTTCTGATCAAGACACATCTGAGACAGAGTTTCAAGAAACA 240
  243 AATTTGTTAAACATCTTATTAAGAACTTTAGTTAGTCTTGAAGTTAAGATTAACAA 302
  241 AATTTGTTAAACATCTTATTAAGAACTTTAGTTAGTCTTGAAGTTAAGATTAACAA 300
  303 AAAAAATTACACAGAGAAACACAAATTAACCACTACCGTCAGTTATCATTAAGATGAA 362
  301 AAAAAATTACACAGAGAAACACAAATTAACCACTACCGTCAGTTATCATTAAGATGAA 360
  363 TGTTTTGTATATTAATTAATTAACACACAAATTAACCTAATTAATTAACATATATGT 422
  361 TGTTTTGTATATTAATTAATTAACACACAAATTAACCTAATTAATTAACATATATGT 420
  423 TATACATATATTTTGTAAAAAATTAGAGTTTTCAAAAACATCTTAATACATGATTAAG 482
  421 TATACATATATTTTGTAAAAAATTAGAGTTTTCAAAAACATCTTAATACATGATTAAG 480
  483 TTTATAGAAATCAATATTTTAAAAATTAATTTTAAAAAATCTTAAAGTCATTA 542
  481 TTTATAGAAATCAATATTTTAAAAATTAATTTTAAAAAATCTTAAAGTCATTA 540
  543 GATCCTCTCACACCTGTGTGATCATTTAGTATGATGATGATGATGATGATGATGAT 602
  541 GATCCTCTCACACCTGTGTGATCATTTAGTATGATGATGATGATGATGATGATGAT 600
  603 AACAGATTAATTAATTAAGATTAACCTGAGAAATATATTAATTAATTAATTAATTA 662
  601 AACAGATTAATTAATTAAGATTAACCTGAGAAATATATTAATTAATTAATTAATTA 660
  663 AAAAAAATCAATTAATTAATTTTGTGTTTCCCATGACATGACAACTGACATGAC 722
  661 AAAAAAATCAATTAATTAATTTTGTGTTTCCCATGACATGACAACTGACATGAC 720
  723 TGCACCTCAGCTCCTCTCTCTCAACACATGTCTCATGTGACATTTGAGCTTTT 782

```


ACCESSION AX463392
 VERSION AX463392.1 GI:21886243
 KEYWORDS
 SOURCE
 ORGANISM Phaseolus vulgaris
 Phaseolus vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Phaseolus.
 REFERENCE 1
 Oulmasov T., Wang Q., Dubois P. and Liang J.
 Arcelin-5 promoter and uses thereof
 Patent: WO 0250295-A 12 27-JUN-2002;
 JOUENAL RENESSEN LLC (US)
 FEATURES
 source
 1.1872
 Location/Qualifiers
 /organism="Phaseolus vulgaris"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3885"
 ORIGIN
 Query Match 69.3%; Score 795; DB 6; Length 1872;
 Best Local Similarity 85.7%; Pred. No. 7.8e-117;
 Matches 1043; Conservative 0; Mismatches 90; Indels 84; Gaps 11;
 Oy 1 TRGATCTCTCAATGAAAGTGTATTTCTCTCATCCAG--ACAAAGGGGCAACGT 58
 Db 645 TAGGATCTCTCAATGAAAGTGTATTTCTCTCATCCAGGAAAGGACCAACGT 704
 Oy 59 TACAAACAAATATGTTTATTTAGATTAAGAGAGTAAAGAGAGAGAGAGATTA 118
 Db 705 TACCAACAAATATTTATTTATTTAGATTAAGAGAGTAAAGAGAGAGAGATTA 764
 Oy 119 AAAAAATCTCTATCTCTTCTT----- 143
 Db 765 AAAAAATCTCTATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 824
 Oy 144 -CTGTAATTAATTAATTAAGAGCTTAACCTTTAATTAATTAATTAATTAAGT 200
 Db 825 ATTGTAATTAATTAATTAAGAGCTTTAATTTAATTAATTAATTAATTAATTA 884
 Oy 201 -----CTAGCATGAGCAGCACTCGAGACAGATTTCAAGAAACATTTGTT 250
 Db 885 TTAGTCAGAGAGTGTATCTTCAACCTTCAGCAACAAATTTCAAGAAACATTTGTT 944
 Oy 251 AAACATCTTTATTAAGAACTTTTAACTCTTGAAGTTAAAGTTAAAGAAATTA 309
 Db 945 AAACATCTTTATTAAGAACTTTTAAAGTTGAAGTTAAAGTTAAAGTTAAAGTT 1004
 Oy 310 ACACAG 369
 Db 1005 ACACAG 1064
 Oy 370 ATATCATTAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
 Db 1065 ATATCATTAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122
 Oy 430 ATATTTTGTAAAGCTTGAAGTTTCAACATTC--TAATACATTAATTAAGTT 488
 Db 1123 ATATTTTGTAAAGCTTGAAGTTTCAACATTC--TAATACATTAATTAAGTT 1182
 Oy 489 GAAATACAAATATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
 Db 1183 GAAATACAAATATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
 Oy 549 CTCACAG 607
 Db 1242 CTCACAG 1301
 Oy 608 ACTAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
 Db 1302 ACTAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1349

Oy 668 GGAATATCAATTAATTAATTTTGAATTTCCCAATGACAGAGAGAGAGAGAG 727
 Db 1350 GGAATATCAATTAATTAATTTTGAATTTCCCAATGACAGAGAGAGAGAGAG 1409
 Oy 728 CCTCAGCTCCCTCTCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 787
 Db 1410 CCTCAGCTCCCTCTCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
 Oy 788 TGACAACTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
 Db 1470 TGACAACTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1529
 Oy 848 CTGGGAG 907
 Db 1530 CTGGGAG 1589
 Oy 908 GCACTAG 967
 Db 1590 GCACTAG 1649
 Oy 968 CCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1027
 Db 1650 CCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1079
 Oy 1028 GCATTCAG 1073
 Db 1710 GCATTCAG 1769
 Oy 1074 AAATTAAG 1131
 Db 1770 AAATTAAG 1829
 Oy 1132 GAGTATGAG 1148
 Db 1830 GAGTATGAG 1846
 RESULT 8
 PHARCLIA 4564 bp DNA linear PLN 27-APR-1993
 LOCUS
 DEFINITION Phaseolus vulgaris arcelin (arc) gene, complete cds.
 ACCESSION M68913
 VERSION M68913.1 GI:169313
 KEYWORDS arcelin; lectin-like seed protein.
 SOURCE
 ORGANISM Phaseolus vulgaris
 Phaseolus vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Phaseolus.
 REFERENCE 1 (bases 1 to 4564)
 Anthony, J.L., Vonder Haar, R.A. and Hall, T.C.
 Nucleotide sequence of a genomic clone encoding arcelin, a
 lectin-like seed protein from Phaseolus
 Plant Physiol. 97, 839-840 (1991)
 JOURNAL
 COMMENT Original source text: Phaseolus vulgaris (Library: Lambda Zap II)
 Young plant leaf DNA.
 FEATURES
 source
 1.4564
 Location/Qualifiers
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /db_xref="taxon:3885"
 /issue_type="leaf"
 /dev_stage="young plant"
 /library="Lambda Zap II"
 gene
 3332..4172
 /gene="arc"
 TATA_signal
 3332..3338
 /gene="arc"
 misc_feature
 3361..3364
 /gene="arc"
 /note="cap region"
 CDS
 3375..4172

```

/gene="arc"
/codon_start=1
/product="arcelin"
/db_xref="GI:169314"
/translation="MSSNLLTALFLVLLTHANSNDASFWETENKNTLLQGDAT
VSEGHLLTNVKNESMGKAFYAPQIDRTIDNLSSTNFTIRINKXNENS
AYGALFALVPGSRPLKGRVYGLFNTTVDADAVAVPDSNRSLIDVNSIRPI
ATESCNFGHNNGEKAVRIYDPSKDLVVSLLYPSSEKCHVASFVPLEKVEDWVS
VGSATSGSKKETTETHNVLSFSSNFIFNGKXKERSNILLNKL"
3375..3437
/gene="arc"
/motif="4169
/gene="arc"
/product="arcelin"

```

ORIGIN

Query Match 23.9%; Score 274.6; DB 8; Length 4564;
 Best Local Similarity 91.6%; Pred. No. 2.5e-34;
 Matches 316; Conservative 0; Mismatches 19; Indels 10; Gaps 2;

```

QY 805 CATGTTGCCACGTAGCTCTCTCTCTCCATGATGACACACTGCGATGATCTGC 864
DB 3028 CACTGACGACCTCAGGCTCTCTCTCTCCATGATGACACACTGCGATGATCTGC 3087
QY 865 CACTGACGCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 924
DB 3088 CACTGACGCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3138
QY 925 CACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 983
DB 3139 CACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3198
QY 984 AAATTAACTAATTATTGATATCTTTTGTATGACGCGATGCGCATTCGCTGT 1043
DB 3199 AAATTAACTAATTATTGATATCTTTTGTATGACGCGATGCGCATTCGCTGT 3258
QY 1044 TTATATAATTGTTAATTGAGTGAATATATAAATAAAGTTGAAAGATTTT 1103
DB 3259 TTATATAATTGTTAATTGAGTGAATATATAAATAAAGTTGAAAGATTTT 3318
QY 1104 GCATTTGTTGTTGTAATAATGAGAGAGATGATGTTATATGCA 1148
DB 3319 GCATTTGTTGTTGTAATAATGAGAGAGATGATGTTATATGCA 3363

```

RESULT 9
 AF193029 2288 bp DNA linear PLN 10-JAN-2000
 LOCUS Phaseolus vulgaris arcelin 5c gene, complete cds.
 DEFINITION AF193029
 ACCESSION AF193029
 VERSION AF193029.1 GI:6684755
 KEYWORDS
 SOURCE Phaseolus vulgaris
 ORGANISM Phaseolus vulgaris
 Buxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Phaseolus
 1 (bases 1 to 2288)
 Gerhardt, I.R. and Grossi de Sa, M.F.
 Molecular characterization of a new arcelin-5 gene
 Unpublished
 2 (bases 1 to 2288)
 Gerhardt, I.R.
 Direct Submission
 Submitted (07-OCT-1999) Embryapa/Cenargen, SAIN-Parque Rural-Final
 W5 Norte, Brasilia, DF 70770-900, Brazil
 Location/Qualifiers
 1..2288
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /cultiivar="G02771"

ORIGIN

Query Match 12.1%; Score 139.4; DB 8; Length 2288;
 Best Local Similarity 85.8%; Pred. No. 8e-13;
 Matches 169; Conservative 0; Mismatches 21; Indels 7; Gaps 1;

```

QY 959 CGCTTCTCCATTAATATCTTTAATTAATTAATTAATTAATTAATTAATTAAT 1018
DB 393 CACTATCTCCATTAATATCTTTAATTAATTAATTAATTAATTAATTAATTAAT 452
QY 1019 GACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
DB 453 GACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 512
QY 1072 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1131
DB 513 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 572
QY 1132 GAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1148
DB 573 GAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 589

```

RESULT 10
 PVPDLEC1 1768 bp DNA linear PLN 12-SEP-1993
 LOCUS P. vulgaris cv. Pinto pseudogene Plect for phytohemagglutinin
 DEFINITION PVPDLEC1
 ACCESSION X04660
 VERSION X04660.1 GI:21020
 KEYWORDS Plect gene; phytohemagglutinin; pseudogene.
 SOURCE Phaseolus vulgaris
 ORGANISM Phaseolus vulgaris
 Buxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Phaseolus
 1 (bases 1 to 1768)
 Voelker, T.A., Staewick, P. and Chrispeels, M.J.
 Molecular analysis of two phytohemagglutinin genes and their
 expression in Phaseolus vulgaris cv. Pinto, a lectin-deficient
 cultivar of the bean
 EMO J. 5, 3075-3082 (1986)
 See also <X04659> For Plect2 gene.
 Several inverted repeats are described in the 5' upstream region of
 the PNA gene.
 None of the 3 possible reading frames allow the translation of its
 message into a complete PNA polypeptide. The PNA translational
 start codon (pos. 708-710) starts a RF which codes for a truncated
 50aa polypeptide with a N-terminal sequence completely identical to
 the dlect gene product, but is out of register after 10 codons due
 to a single bp deletion (pos. 739/740).
 Location/Qualifiers
 1..1768
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /strain="Pinto U111"

```

/db xref="taxon:3889"
/i:sue.type="young leaves (greenhouse)."
promoter      665..670
              /note=put. TATA-box"
CDS           708..865
              /note="unnamed protein product; pseudogene region"
              /codon_start=1
              /protein_id="CAA28363.1"
              /db_xref="gi:21021"
              /db_xref="EMBL:CAA28363"
              /translation="MASSNLSLALSLSPSTQTPAKPSASISGRKPLTSSKMP
SHPASYD"
variation     739..740
              /note="tct in Pdic2 (c deletion, causing a frame
misc_feature   1520..1635
              shift mutation)"
              /note="put. polya signal"

ORIGIN
Query Match          10.3%; Score 118.2; DB 8; Length 1768;
Best Local Similarity 64.4%; Pred. No. 1.9e-09;
Matches 352; Conservative 0; Mismatches 125; Indels 70; Gaps 9;

OY    636 ATATATTAATTATATCATTTAAATTAAGGAAAATCAATTAGAATTTTGATGCC 695
Db    186 ATTATTAATTATTAATTAATTAATTATTAACCAACGGGTCTMGACTGGCGGGT 245
OY    696 CCACATGCACAACATCACATGCACGCGCCACCCTCAGCTCCCTCTTCACACATGC 755
Db    246 TNACATGCACACATGCACATGCANSCGACCCCTCAGCTCCCTCTTCACAGATC 305
OY    756 -TCATGTCACTTTCAGCTTTGGCTTTTCACATAGACGAACCTGCCATGCAT---GT 810
Db    306 TTCCTCTCACTTCACATGCACGCTTTCACACATAGACATCAGTGTCAATGACAGCT 365
OY    811 GCCACGGAGGCTCGTCCCTCTTCCCATATGACACACACTGGGACATGCATGCACCTC 870
Db    366 GCACCTCAGCTTTCTCCTCTTTCTCAATGACAGACATGGCCATGCATGANGA----- 421
OY    871 AGCTCCACCTCTTCTCAATTAGAGCTACTGCCCCATGCACCTGCAGACCTC 930
Db    422 TGCTGGCACCCTCAGCTC-----CTCTCTTTGAC 450
OY    931 TCCTACTTCCTCATGTCTACCGCCAACCGCTCTCTCATTAATATCATTTAAATTGA 990
Db    451 CCGTGTCTCCCAAGCTACGATGCC-AACCGCTCTCTATATACATATCTTTAAATTGA 509
OY    991 AA-CATAATATTTCAAT------ACTTTTTGATGACGTGATGCATTCGACATGTTG 1047
Db    510 AACCTAATTTATTTCAATTTTTAAATGTTTTGATGACGTGATGCATTCGACATGTTG 569
OY    1043 TTATTAATATTTGTTAATTTTGACATT-----GATTAATTAATGAAAAAAGTTG- 1092
Db    570 CTAAATCTTAATTTCTATATTTCTTTATCTTCAATTAATATATATAAGAAAATCAT 629
OY    1094 -----GANAAGTTTTGCATTTGTGTGTATTAATATAGAGAGAGATGATGTT 1141
Db    630 GAAGTAGAGAGAGAGAGAGATGCATTGTGTGTATTAATTAAGAGAGATGATGTT 689
OY    1142 TAATGCA 1148
Db    690 TAATGCA 696

RESULT 11
PEAVAR23A    PFVAR23A       22243 bp  DNA        linear INV 14-SRP-1591
LOCUS        Plasmodium falciparum (strain FCR3) variant-specific surface
DEFINITION   Protein (var-2, var-3) genes, complete cds's.
ACCESSION    LA0609
VERSION      LA0609.1 GI:886376
KEYWORDS     Variant-specific surface protein.
SOURCE       Plasmodium falciparum [malaria parasite P. falciparum]

```

[illegible]

REFERENCE 1 (bases 1 to 1395)
 AUTHORS Hoffman, L.M. and Donaldson, D.D.
 TITLE Characterization of two Phaseolus vulgaris phytohemagglutinin genes closely linked on the chromosome
 JOURNAL EMBO J. 4 (4), 883-889 (1985)
 MEDLINE 85257505
 PUBMED 2990911
 COMMENT Data kindly reviewed (06-MAR-1986) by D.D. Donaldson.
 FEATURES
 source location/Qualifiers
 1..1395
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /db_xref="taxon:3885"
 repeat_unit 10..29
 /note="1mp, inverted repeat A"
 12..30
 /note="RY repeat (alternating pyrimidines and purines)"
 repeat_region 12..26
 /note="direct repeat 1"
 repeat_unit 74..93
 /note="1mp, inverted repeat A"
 74..91
 /note="RY repeat"
 misc_feature 76..90
 /note="direct repeat 1"
 269..273
 /note="pot. 'CAAT' element"
 332..337
 /note="pot. TATA box"
 360..364
 /note="cap region"
 367..377
 /note="cap repeat"
 375..1202
 /codon_start=1
 /product="erythroagglutinating phytohemagglutinin"
 /protein_id="CAA26256.1"
 /db_xref="GI:758251"
 /db_xref="GOA:P05088"
 /db_xref="MISC-PROT:P05088"
 /translation="MSSNLISLALFLVLTTHANSAGTSFSFORFNETNLIORDAT
 VSSKQRLTNVNDGEPTLSLGRAPFSAPIQIWDNTTGVAASPTSFPTNIDVFN
 SGPDAGLAFVLLPVGSGPKDKGLGLFNNTKYSNATVAVEPTLYNVMDPKPRH
 IGIDVNSIKSIKTTTWDPVKENAEVLITTYOSTLVASLIVPSLKTSPVSDIVDL
 KSLPEWYVIGFTATGILTKGNVETNDILMSFSKISDGTTSBALNANALNQIL"
 375..437
 /note="signal peptide (aa -21 to -1)"
 438..1199
 /product="erythroagglutinating phytohemagglutinin (aa
 1-254)"
 471..479
 /note="glycosylation site"
 615..623
 /note="glycosylation site"
 675..683
 /note="glycosylation site"
 923..930
 /note="inverted repeat B"
 939..944
 /note="pot. transposable element insertion site"
 949..956
 /note="inverted repeat B"
 1294..1299
 /note="polyadenylation signal"
 1298..1304
 /note="polyadenylation signal"
 ORIGIN

Query Match 9.4%; Score 108.4; DB 8; Length 1395;
 Best Local Similarity 69.8%; Freq. No. 7.2e-08;
 Matches 240; Conservative 0; Mismatches 56; Indels 48; Gaps 5;
 852 GCATGACGCTGCACCTGACCTCCACCTCTTCATTAAGAGCCTACTGGCATGCA- 910

Db 21 GCATGACGCTGCACACAGCTTCTCTCTTTCACTATGACAGCACTGGCCATGCA 80
 Oy 911 -----CACTGCCACTGCAGACTCTCTACCTCCCATTTGCCATTCACCTGCC 954
 Db 81 GCATGACGCGGCGACCTGACCTCCCTCTCTACCCGCTTCCAAATGACGCGCC 140
 Oy 955 AACCCTCTCTCCATTAATATCTATTAAATTTAACTAACTATTATTTACTTT-- 1012
 Db 141 -AACCGCTCTCTCTAATATCTCTTTAAATTTAACTAACTATTATTTACTTTCA 199
 Oy 1013 -----TTGATGACGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1056
 Db 200 ATGATTTGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 259
 Oy 1057 ATTGAGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1104
 Db 260 TTACTCCCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 319
 Oy 1105 CATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1148
 Db 320 CAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 363
 RESULT 13
 PHVDLECA 1441 bp DNA linear PLN 27-APR-1993
 LOCUS P. vulgaris phytohemagglutinin gene encoding erythroagglutinating
 DEFINITION phytohemagglutinin (PHA-E), complete cds.
 ACCESSION K03288
 VERSION K03288.1 GI:169336
 KEYWORDS haemagglutinin; lectin; phytohemagglutinin.
 SOURCE Phaseolus vulgaris
 ORGANISM Phaseolus vulgaris
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Phaseolus.
 1 (bases 1 to 1441)
 Hoffmann, L.M.
 JOURNAL Unpublished (1985)
 REFERENCE 2 (bases 1 to 1395)
 AUTHORS Hoffman, L.M. and Donaldson, D.D.
 TITLE Characterization of two Phaseolus vulgaris phytohemagglutinin genes closely linked on the chromosome
 JOURNAL EMBO J. 4 (4), 883-889 (1985)
 MEDLINE 85257505
 PUBMED 2990911
 COMMENT Original source text: P. vulgaris cv. Tendergreen DNA, clone
 lambda-B10.
 Two tandem polyadenylation signals are present at positions
 1294-1299 and 1299-1304. Direct and inverted repeats in the form
 of alternating pyrimidine and purine residues are located at
 positions 10-29 and 74-93, and an two inverted repeats were found
 at positions 923-930 and 949-956.
 Draft entry and sequence [1] in computer-readable form were kindly
 provided by L.M. Hoffman (07-OCT-1985).
 FEATURES
 source location/Qualifiers
 1..1441
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /db_xref="taxon:3885"
 360..31441
 /product="d1ec1 mRNA"
 375..1202
 /note="phytohemagglutinin prepeptide"
 /codon_start=1
 /protein_id="AAA33759.1"
 /db_xref="GI:169337"
 /translation="MSSNLISLALFLVLTTHANSAGTSFSFORFNETNLIORDAT
 VSSKQRLTNVNDGEPTLSLGRAPFSAPIQIWDNTTGVAASPTSFPTNIDVFN
 SGPDAGLAFVLLPVGSGPKDKGLGLFNNTKYSNATVAVEPTLYNVMDPKPRH
 IGIDVNSIKSIKTTTWDPVKENAEVLITTYOSTLVASLIVPSLKTSPVSDIVDL

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:08:39 ; Search time 3381 Seconds
(without alignments)
10139.540 Million cell updates/sec

Title: US-10-015-637-1
Perfect score: 1148
Sequence: 1 taggactctcaatagaaaa.....agagatgacgttaacga 1148

Scoring table: IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: em_estba:*
10: em_estba:*
11: em_estba:*
12: em_estba:*
13: em_estba:*
14: em_estba:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: em_estba:*
29: em_estba:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117.8	10.3	1101	29	CNS00EVL
2	104.6	9.1	1210	29	CG749728 P044-1-C0
3	103.2	9.0	1201	9	AL536104
4	101.8	8.9	1201	9	AL565455

removed
of the
16 gene

C 5	99.6	8.7	1191	29	CG754863	CG754863 P050-2-C0
C 6	99.4	8.7	1200	13	BX415878	BX415878 BX415878
C 7	98	8.5	1201	13	BX439779	BX439779 BX439779
C 8	97.4	8.5	614	29	CNS0152H	AL104915 Drosophila
C 9	97.4	8.5	1193	29	CG745316	CG745316 P038-1-G0
C 10	97	8.4	1201	13	BX462207	BX462207 BX462207
C 11	96.8	8.4	1200	13	BX437758	BX437758 BX437758
C 12	96.6	8.4	1200	13	BX436885	BX436885 BX436885
C 13	96	8.4	620	28	BH183498	BH183498 023 L 07-
C 14	96	8.4	620	28	BH183498	AL520449 03 L07 OF
C 15	95.8	8.3	1348	29	CG749499	CG749499 P043-4-A0
C 16	94.6	8.2	932	9	AL514901	AL514901 AL514901
C 17	93.6	8.2	1201	13	BX443774	BX443774 BX443774
C 18	93.2	8.1	1201	29	CNS01JRG	AL147405 Anopheles
C 19	93.2	8.1	1201	29	CNS0167M	AL106396 Drosophila
C 20	93.2	8.1	1056	13	BX415058	BX415058 BX415058
C 21	92.8	8.1	1200	13	BX437739	BX437739 BX437739
C 22	92.8	8.1	1200	13	BX415636	BX415636 OR20714 M
C 23	92.4	8.0	1092	29	CNS020K7	AL115636 Tetradodon
C 24	92.4	8.0	1101	29	CNS0039G	AL106392 Drosophila
C 25	92	8.0	1190	29	CNS020M7	AL206908 Tetradodon
C 26	91.8	8.0	1048	13	BX456942	BX456942 BX456942
C 27	91.4	8.0	1201	13	BX420717	BX420717 BX420717
C 28	91.4	8.0	1201	13	BX462660	BX462660 BX462660
C 29	91.4	8.0	1201	13	BX462660	CG702001 P044-3-G0
C 30	91.2	7.9	1175	29	CG750201	CG750201 P052-4-C0
C 31	91	7.9	1392	29	CG757503	CG757503 P052-4-C0
C 32	90.6	7.9	1646	14	CF547155	CF547155 AGNCOURT
C 33	90.6	7.9	1201	9	AL536104	AL536104 AL536104
C 34	90.4	7.9	996	29	CNS000F0H	AL071063 Drosophila
C 35	90.4	7.9	1201	13	BX443774	BX443774 BX443774
C 36	90.4	7.9	1531	29	CG748014	CG748014 P041-4-B0
C 37	90.2	7.9	1135	29	CNS033CQ	AL226115 Tetradodon
C 38	90	7.8	1566	29	CG757757	CG757757 P053-1-D0
C 39	89.8	7.8	1626	14	CF238805	CF238805 AGNCOURT
C 40	89.6	7.8	1095	28	CC264029	CC264029 CH261-57A
C 41	89.6	7.8	1201	13	BX361134	BX361134 BX361134
C 42	89.6	7.8	1364	29	CG757966	CG757966 P053-2-D0
C 43	89.4	7.8	895	29	CNS06FEV	AL396821 T7 end of
C 44	89.4	7.8	1098	13	BX377526	BX377526 BX377526
C 45	89.4	7.8	1200	13	BX415878	BX415878 BX415878

ALIGNMENTS

RESULT 1
LOCUS CNS00EVL
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29823 of RPL1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706.1 GI:494849
VERSION AL069706
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Bukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygora;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope:
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Oseguwa and
Aaron Mammosser in Pleter de Jong's laboratory in the Department of

QY 493 TACAAATTTTAAAAATATATTTTAAAAACATTTCTAAAGTCATTGAGATCTCTCA 552
 DB 321 NANN 262
 QY 553 CACCTGTGTGATCTTTAGTCATGTATGTACATCATCTTACTTGTACAGAGATTA 612
 DB 261 AA 202
 QY 613 AATAAATAGATTAAGGAT 672
 DB 201 AA 142
 QY 673 ATCAATTTAGAAAT 686
 DB 141 AAAAAAAAAAAATTT 128

RESULT 3
 AL536104/c 1201 bp mRNA linear EST 31-MAY-2003
 LOCUS
 DEFINITION CS0DF022YC18 5-PRIME, mRNA sequence.
 ACCESSION AL536104.2 GI:31260974
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12939597.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DF022BB09QPL.

FEATURES
 source
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DF022YC18"
 /tissue_type="FETAL BRAIN"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL BRAIN"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 9.0%; Score 103.2; DB 9; Length 1201;
 Best Local Similarity 39.7%; Pred. No. 2.3e-08;
 Matches 173; Conservative 85; Mismatches 171; Indels 7; Gaps 1;
 QY 89 TTAAGAGGTAAG 148
 DB 1085 YTAAG 1026
 QY 149 ATAAAT 208
 DB 1025 WAAAG 966
 QY 209 GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 268

DB 965 ATTAATTCAT 906
 QY 269 TTTAGTAAAGCTCTGAGGTAGATTAACAAAAAATTTACACAGAGAGAGAGAGAG 328
 DB 905 TATATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 846
 QY 329 AACCCAGTCAGCTGAGGTATCATPAGAGATGAATGTTTGAATTCATTAATTAACAC 386
 DB 845 AA 789
 QY 389 ACACAAAAATACATCTATATATTAACAAATATATATATATATATATATATATATAT 448
 DB 788 -----WAT 733
 QY 449 GAGTTTTCAGAAATCTAT 508
 DB 732 AAT 673
 QY 509 AT 524
 DB 672 HAHTAT 657

RESULT 4
 AL565455 1201 bp mRNA linear EST 12-MAY-2003
 LOCUS
 DEFINITION CS0DF005Y018 3-PRIME, mRNA sequence.
 ACCESSION AL565455
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 16, 2001 this sequence version replaced gi:12916848.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 9232.f for
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DF005BH09NP1&cluster=9232.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DF005BH09NP1.

FEATURES
 source
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DF005Y018"
 /tissue_type="FETAL BRAIN"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL BRAIN"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 8.9%; Score 101.8; DB 9; Length 1201;
 Best Local Similarity 37.0%; Pred. No. 4.1e-08;
 Matches 228; Conservative 115; Mismatches 287; Indels 3; Gaps 3;
 QY 53 AACAGTTAAG 112

[illegible]

```

source
1..1391
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa Ecoli BAC Library"
/notes="The library was generated by a partial digest of the genomic DNA with Ecoli and cloning into the BAC vector."

ORIGIN
Query Match      8.7%; Score 99.6; DB 29; Length 1391;
Best Local Similarity 48.7%; Pred. No. 9.3e-08;
Matches 329; Conservative 0; Mismatch 3; Indels 8; Gaps 3;

QY 12 AATGAGAAAGTGTATTTCTCATCATCACAGACACAAAGGGCAACAGTTAAACAAACAAT 71
Db 1015 AATTATTAATATATATTTTAAATATAATATAAATAATTATATATAATTTTAAATATTAAT 956
QY 72 TTATGTTTCATTTGAGATTAAGGAAGTAAGGAAGAAAGAAAGAAAGATTAAGAAATATGTCCTT 131
Db 955 TAAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 896
QY 132 A--TCCTCTGTTCTGTATATAATATAATAGAGCTTAACTTTAATATAATATG 188
Db 895 AATTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 836
QY 189 TAATTAGCTTTCTAGTCATGAGACACACTCAGACAGACAGATTCAGAGAAACAATTTG 248
Db 835 AATTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 776
QY 249 TTAAACATCTTATTTAGAACTTTTGTAACTTTGAAAGTTAGATTTAAAGAAAAAT 308
Db 775 TTTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 716
QY 309 TACACAGAGAGACACACATTAACCCACTACCGGTATCATAGAGATGAATGTTT 368
Db 715 TATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 656
QY 369 GATATCATTAATATATACACACACAAATATCATCTTAATTTAAACATATATGTATACA 428
Db 655 AATTAATATATATATATATATATATATATATATATATATATATATATATATATAT 597
QY 429 TATATTTTGTAAACCTTAGAGTTTTCAAAACATCTTAATATCATATATAGATTTA 488
Db 596 TTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 541
QY 489 GAATATACAAATATTTAAAAAATATATATTTTAAAAAATCATTAAAGTCATTCAGATCT 548
Db 540 TAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 481
QY 549 CTCACACTCGTGTATCATTTAGCATGTATGTAGTCAATCATTTGTGTGTACACACAGA 608
Db 480 TAAAAATATATATATATATATATATATATATATATATATATATATATATATATAT 421
QY 609 GTAAATATTAATATAGATTAATCTAGAGATATATATATATATATATATATATATATAT 668
Db 420 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 361
QY 669 GAAATATCAATTAGAA 684
Db 360 AAAAAAAAAAAAAAAAAAAAA 345

RESULT 6
LOCUS BK415878/c 1200 bp mRNA linear EST 15-MAY-2003
DEFINITION BK415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008Y104
ACCESSION BK415878
VERSION BK415878.1 GI:30765550
KEYWORDS Bst.
SOURCE Homo sapiens (human)

```


ORIGIN vector "

Query Match 8.5%; Score 97.4; DB 29; Length 1193;
Best Local Similarity 46.5%; Pred. No. 2.4e-07;
Matches 313; Conservative 0; Mismatches 357; Indels 3; Gaps 2;

43 ACAAAGGCGACAGCTTACCAAAACAATTATGTTGATTTGAGATTAAGAGAGTAA 102
1024 AAAAAAT 965
103 GAGAGAAAAAGATTA-AAAAATGCTCTTCTCTTCTCTCTCTCTCTCTCTCTCT 161
964 AT 905
162 AGACTTAACTTTTAAATATATATATATATATATATATATATATATATATATAT 221
904 AT 845
222 AGACAGATTTCAAGAAACAATTGTTAAACATCTATTAGAACTTTAGTTAAGTC 281
844 TATATAAT 785
282 TTGAAGTTAGATTTAAACAAAAATTTACACAGCAGAAACACATTAACCCACT 341
784 TAT 725
342 CAGGTTATCATAGAGATGAATGTTTATATCATTAATTAATTAACACACAAAAAT 401
724 ATTAAAAAAAT 665
402 TCTAAT 461
664 TAT 605
462 CATCTTAATCATGATGATGATTTTACGAATCAAAATTTTAAATATATATATTA 521
604 ATNAAAAAT 547
522 AAAACATCTTAAAGCATTCAGATCCTCTCACACCTGTGTGATCATTAGCATAT 581
546 AT 487
582 AGTACATCATGTTGATTCACACAGAGTAAATATATATATATATATATATATAT 641
486 AT 427
642 AT 701
426 AAAAAAT 367
702 GACACACTCACCC 714
366 NNNNNNNNNCCCC 354

RESULT 10
BX462207/c 1201 bp mRNA linear EST 22-MAY-2003
LOCUS BX462207 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
DEFINITION clone CS0D004YB03 5-PRIME, mRNA sequence.

ACCESSION BX462207.1 GI:31023422
VERSION BX462207.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Poljansky, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 24.r For more
information about this cluster, see http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D004CA020piccluster=24.r. Contact :
Peng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D004CA020P1.

FEATURES

source
1..1201
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D004YB03"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/note="Vector: PCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the PCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match 8.4%; Score 97; DB 13; Length 1201;
Best Local Similarity 42.6%; Pred. No. 2.8e-07;
Matches 253; Conservative 50; Mismatches 287; Indels 4; Gaps 1;

100 AAGGAAAGAAAAAGATTAATAAAATGCTTTATCTCTTGTTCGTGATTAATATATA 159
1135 AAAAAAT 1076
160 AGACACTTAACTTTTATATATATATATATATATATATATATATATATATATAT 219
1075 AAAAAAT 1016
220 AGACAGAGATTTCAAGAAACAATTGTTAAACATCTTATAGAACTTTTGTAAAG 279
1015 AAAAAAT 957
280 TCTTGAAGTTAGATTTAAACAAAAATTTACACAGCAGAAACCAATTAACCACT 339
956 ---AAAAAT 900
340 GTCAGTTATCATAGAGATGAATGTTTGTATCATTAATATATATATATATATAT 399
899 TTTTAAAAAT 840
400 CATCTAATTTTAAACATATATATATATATATATATATATATATATATATATAT 459
839 AAAAAAT 780
460 AACATCTTAAATCATGATTAAGATTTATAGAAATACATATATATATATATATAT 519
779 AAAAAAT 720
520 AAAAAAT 579
719 TAAAAAT 660
580 GTAGTACATCATGTTGATTCACACAGAGTAAATATATATATATATATATATATAT 639
659 TMTTAAAAATTTTMTTAAAAATATATATATATATATATATATATATATATATAT 600
640 AT 693
599 AAAAAAT 546

RESULT 11
BX437758

	LOCUS	BK437758	1200 bp	mRNA	linear	EST 15-MAY-2003
	DEFINITION	BK437758 Homo sapiens THYMUS Homo sapiens CDNA clone CSOCAP008XB01				
	ACCESSION	BK437758				
	VERSION	BK437758.1	GI:30773605			
	KEYWORDS	EST.				
	SOURCE	Homo sapiens (human)				
	ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	REFERENCE	1 (bases 1 to 1200) Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished (2001)				
	AUTHORS	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paradise Avenue Genoscope sequence ID : CSOCAP008CA010P1.				
	TITLE	JOURNAL				
	JOURNAL					
	COMMENT					
	FEATURES					
	source	1..1200 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSOCAP008XB01" /tissue_type="THYMUS" /clone_lib="Homo sapiens THYMUS" /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a Molt-oliigo(dT) primer. Iste prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector." Library was not normalized."				
	ORIGIN					
	Query Match	8.4%; Score 96.8; DB 13; Length 1200;				
	Best Local Similarity	35.4%; Pred No. 3e-07;				
	Matches	228; Conservative 120; Mismatches 291; Indels 5; Gaps 2				
D	53 AACAGTTAACAAACAATTTAATGTTTCATTGAGATTAGAGAAGTAAGAGAAA	112				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	553 AAAAAATATAATTTTTTTWMAAAAAASGAARAKATTTTTTKTTTATWTWKRGATWTYT	612				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	113 GATTAAAAAATGTCCTATCTCTTGTTCTGTGAATTAATTAATTAAGACTTAAC	172				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	613 TTTTWTWMAASDDMTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWRDARRTAKATKA	672				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	173 TTAAATATTAATATGTATATTAGTTTCTAGTCAAGACACACTCAGAGACAAGATT	232				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	673 WTTTWMTTAMAWAGGAARRRRAATWTWTWMAAAMWAGAAAGAGAGAAATTTTTTTA	732				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	223 CAAGAAAACAATT---TTGTTAAACATCTTTTGTGAAACCTTTTAGCTTGAAGT	288				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	723 TTAGARGATWTTTAAWMTATATARGARGADTTTTTATATMTATTTTTTTWMAAGATDK	792				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	289 TAGAATTTAAACAAAAAATTTACACAGCAAGAACCAATTAACCACATCCGTCAAGTGA	348				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	793 AAAAAWWMTWTWTAAAAAATTTWWWAGABAARAKTWTWTWAAAAAAMBARWAAT	852				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	349 TCATTAAGCATGAATGTTTGATATCATTAATTTATACACACAAAAATTAAGCTAAT	408				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	853 ATATTTTATATATATKABAAAAAATTAATAAARARERKGAIAAAAAATTAATAAT	912				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	409 ATAACAATATATGTTATATACATATTTTTTGTAAAAACCTTAGAGITTTTCAAACAATCTA	468				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	913 TTTATATAAAAAATTTTAAWMBRAABARAAMWAAAAAGAAARAGAAWAAATTTTTTTA	972				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	469 ATACATGATTAGAGTTTATGAAATACAAATTTTAAAAAATATTAATTTTAAAAAACAAT	528				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
D	973 ATATATAAAAAAMWTTTTTTTTTWTMAAAAAAMWDTTTTTATAT--WTATATATGAAATAAT	103				
O	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					

Oy		529	TCCAAAGTCATCTCAGATCCCTCACACCCTGTGACCATTTAATCACTGTAATGAGTACA	588
Dd		1032	AACATTTAAATTATWADAAAARAAAGAATATDAGAGATAAAATTTTTTTTTTTTTTTAAAA	1091
Oy		589	TCATTTGGATTCACACAGATTAATAATTAATAGATTAACCTAGGAATATATATATAT	648
Dd		1092	AAAGAAAAATTTATARAFAAAAAAAAAAWWWTAAAAAAAAAAAAAAAAAAAAAGAACDAWAA	1151
Oy		649	ATACAAATTAATTAATAAAGGAATCAATTAATTTGATTTGAT	692
Dd		1152	GAAATATTAATAAARAAAAAGTAAARAAARAAAAATTAATAT	1195
<hr/>				
RESULT 12				
BX436885 998 bp mRNA linear EST 15-MAY-2003				
LOCUS BX436885				
DEFINITION BX436885 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP004YC09				
ACCESSION BX436885				
VERSION 5-PRIME, mRNA sequence.				
KEYWORDS BX436885.1 GI:30785562				
SOURCE EST.				
ORGANISM Homo sapiens (human)				
REFERENCE Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE L4.W.B., Gruber,C., Jesse,J. and Polayes,D.				
JOURNAL Full-length cDNA libraries and normalization				
COMMENT Unpublished (2001)				
Contact: Genoscope				
Genoscope - Centre National de Sequencage				
BP 191 91006 Evry cedex - France				
Email: seq@genoscope.cns.fr, web : www.genoscope.cns.fr				
Library was constructed by life technologies, a division of				
Invitrogen. Contact : Feng Liang Email: fliang@life tech.com URL :				
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600				
Faraday Avenue Genoscope sequence ID : CSOCAP004B05OP1.				
<hr/>				
FEATURES				
Source				
1..998				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref="taxon:9606"				
/clone="CSOCAP004YC09"				
/tissue_type="THYMUS"				
/clone_id="Homo sapiens THYMUS"				
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed				
with a NotI-oligo (dT) primer. Five prime end enriched,				
double-strand cDNA was digested with Not I and cloned				
the Not I and EcoRV sites of the pCMVSPORT 6 vector.				
Library was not normalized."				
<hr/>				
ORIGIN				
Query Match 8.4%; Score 96.6; DB 13; Length 998;				
Best Local Similarity 43.1%; Pred. No. 3,4e-07;				
Matches 283; Conservative 39; Mismatches 331; Indels 3; Gaps 1;				
Oy		53	AACGCTAACAAACAAATTTATCTTCATTGAGTTAAGAGAGTAAGAGAAAAAA	112
Dd		288	AMTATATATATATTTTAAATATATTAATTTATTTATTAATAATATATATTAATAAATAA	347
Oy		113	GATTAATAAANAATGCTTATCTCTTGTTCTGTAAATATATATATATATAGACTTAACT	172
Dd		348	AATTCAAAAAAGAAAAAATATTAATAAATAAATAAATAAATAAATAAATAATWTAAATTAAA	407
Oy		173	TTTAAATATATATTTGTAATTTAGTTTTCAGTGCATAGACGCCACTGACAGACAAGATT	232
Dd		408	TTTAAAT	467
Oy		233	CAGAAAAACAATTTGTTGAACATCTTATTTGAAACTTTTAGTTAGCTTTGAAGTTAGA	292
Dd		468	AAAAAATAAT	527
Oy		293	ATTAAACAAAAAATTTACACAGAGAAACACATTAACCACTACCGTCGATATATAT	352


```
Oy 457 CAAAACATTCATACATGATTAGAGTTATAGAAATACAAATATTTAAAAATTAATT 516
    |||||
Db 424 AAAAAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 365
    |||||
Oy 517 TTAATAAAACATTTTAAAGTCATTCAGATCCTCCACACCTGTGATCAATTAGTCATG 576
    |||||
Db 364 TAAATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 305
    |||||
Oy 577 TATGTAGTACATTCATGTAGTTCACACAGTAAATTAATTAAGATTAACCTAGGAA 636
    |||||
Db 304 AAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 245
    |||||
Oy 637 TATATTAATATATACAAATTAATAATAATAATAATAATAATAATAATAATA 684
    |||||
Db 244 ATTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 197
    |||||
```

Search completed: June 2, 2004, 03:32:47
Job time : 3393 secs

Wed Jun 2 16:36:20 2004

us-10-015-637-1.rml

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 01:03:24 ; Search time 108 Seconds
(without alignments)
5898.922 Million cell updates/sec

Title: US-10-015-637-1

Sequence: 1 tagagctcctcaatagaagaa.....agagagctgagtgatgca 1148

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCFUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	6.4	19124	2	US-08-487-826B-13
2	72.2	6.3	7218	1	US-08-232-463-14
3	72	6.3	5562	4	US-10-204-708-63
4	71.8	6.3	6866	4	US-10-204-708-20
5	70.6	6.1	837	3	US-08-998-416-288
6	68.2	5.9	6669	3	US-10-204-708-6
7	67	5.8	636	3	US-08-998-416-1137
8	66.6	5.7	8093	4	US-10-204-708-32
9	65.8	5.7	10467	2	US-08-487-826B-13
10	65.8	5.7	19124	2	US-08-487-826B-13
11	64.8	5.6	6306	4	US-10-204-708-50
12	64.6	5.6	6317	4	US-10-204-708-12
13	64	5.6	5610	4	US-10-204-708-54
14	63.8	5.6	660	1	US-07-991-867B-32
15	63.8	5.6	660	1	US-08-107-755A-32
16	63.8	5.6	660	2	US-08-544-332-32
17	63.8	5.6	660	2	US-09-370-861A-32
18	63.8	5.6	1511	1	US-07-991-867B-8
19	63.8	5.6	1511	1	US-08-107-755A-8
20	63.8	5.6	1511	2	US-08-544-332-8
21	63.8	5.6	1511	2	US-09-370-861A-8
22	63.8	5.6	4810	3	US-08-852-629-11
23	63.8	5.6	4838	3	US-08-852-629-15
24	63.8	5.6	6040	4	US-10-204-708-69
25	62.6	5.5	6124	4	US-08-213-419B-3
26	62.4	5.4	6156	4	US-10-204-708-60
27	62.2	5.4	8961	4	US-10-204-708-80

c 28	61.4	5.3	640681	4	US-09-790-988-1	Sequence 1, Appl
c 29	61.2	5.3	6317	4	US-10-204-708-11	Sequence 11, Appl
c 30	60.6	5.3	665	2	US-08-883-795A-36	Sequence 36, Appl
c 31	60.2	5.2	1501	4	US-09-457-037B-30	Sequence 30, Appl
c 32	60.2	5.2	1501	4	US-09-733-151-30	Sequence 30, Appl
c 33	59.8	5.2	6801	4	US-10-204-708-61	Sequence 61, Appl
c 34	59.4	5.2	694	4	US-09-457-037B-39	Sequence 39, Appl
c 35	59.4	5.2	694	4	US-09-733-151-39	Sequence 39, Appl
c 36	59.4	5.2	1279	4	US-09-457-037B-40	Sequence 40, Appl
c 37	59.4	5.2	1279	4	US-09-733-151-40	Sequence 40, Appl
c 38	59.4	5.2	640681	4	US-09-790-988-1	Sequence 1, Appl
c 39	59.2	5.2	5610	4	US-08-213-419B-3	Sequence 3, Appl
c 40	59.2	5.2	6124	4	US-10-204-708-24	Sequence 24, Appl
c 41	59	5.1	11049	4	US-10-204-708-24	Sequence 3, Appl
c 42	59	5.1	53332	4	US-09-801-861-3	Sequence 186, App
c 43	58.6	5.1	615	3	US-08-998-416-186	Sequence 13, Appl
c 44	58.6	5.1	6113	4	US-10-204-708-13	Sequence 4, Appl
c 45	58.2	5.1	1218	2	US-08-731-722-4	

ALIGNMENTS

RESULT 1
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSER: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH21.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-6550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13
Query Match 6.4%; Score 74; DB 2; Length 19124;
Best Local Similarity 48.3%; Pred. No. 2.8e-07;

Wed Jun 2 16:36:20 2004

us-10-015-637-1.rnd

Page 2

Matches 304; Conservative 0; Mismatches 315; Indels 11; Gaps 3;

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

1 PRIOR APPLICATION DATA: US/07/935,313
2 APPLICATION NUMBER: 05
3 FILING DATE:
4 APPLICATION NUMBER: EP 91 114 300.6
5 FILING DATE: 26-AUG-1991
6 ATTORNEY/AGENT INFORMATION:
7 NAME: SANTI, Stephen A.
8 REGISTRATION NUMBER: 29,768
9 REFERENCE/DOCKET NUMBER: 30472/114 IMPDET
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (703) 836-9300
12 TELEFAX: (703) 683-4109
13 TELEX: 899149
14 INFORMATION FOR SEQ ID NO: 14:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 7218 base pairs
17 TYPE: nucleic acid
18 STRANDEDNESS: single
19 TOPOLOGY: linear
20 IMMEDIATE SOURCE:
21 CLONE: pTZ57-Fla
22 US-08-232-463-14

```

US-08-232-463-14

Query Match 6.3%; Score 72.2; DB 1; Length 7218;

Matches	23;	Conservative	204;	Mismatches	122;	Indels	0;	Gaps	0;
---------	-----	--------------	------	------------	------	--------	----	------	----

```

Qy      685  TTTTGTATCCCAATACACAACTACACATGACGCTGACACCTGACCTCTCT
      1127  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Qy      745  CCACACAGTCTCAGATGACCTTGAGACTTTGACCTTTTACATATGACCAACTCGGAG
      1187  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db      805  CAGTGTGCACAGTACGCTCTCTCTCCATGATGACACCACTGGAGATGACGTGC
      1247  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Qy      865  CACTTCAGTCCCACTCTTCTCATATATGAGCTATGCGCATGCACTGCACTCG
      1307  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db      925  CACTCCTCATCTTCCATGCTACTGCAACCGCTCTCCATAATATCTATTTA
      1367  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Qy      985  AATTATTAATAATTTCATATATCTTTTGTAGATGAGTGAAGCATTTG 1033
      1427  YYYYYYGTACCAATCTCTCATCTCTTAACTACTGACATGATAG 1475
Db

```

RESULT 3 -
US-10-204-708-63/C
Sequence 63, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204.708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 63
LENGTH: 5562
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

Query Match 6.3%; Score 72; DB 4; Length 5562;
Best Local Similarity 47.4%; Pred. No. 5.2e-07;
Matches 216; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

QY 253 ACATCTTATAGAACTTTAGTGAAGTCTTGAAGTGAATTAACAAAAAATTAACA 312
DB 1864 ACTTCAATTAATAATACTTAATACTTAATAATAATAATACTTAATAATA 1805
QY 313 CAGAGAAACATTAACCCCTACCGTACGTTATCATAGAGTAATGTTTGATA 372
DB 1804 AAAAAAATCTACCATCTAATAAATAATCATCTAATCTCAAAACGAAACCTAATA 1745
QY 373 TCATTAATATACACACACACAAATCATCTAATATATACATATATATACATATA 432
DB 1744 AATTCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1685
QY 433 TTTTGTAAACCTTAGAGTTTTCAAACATTCATTAATACATAGTTAGTTATAGAA 492
DB 1684 ATATATATATTAATAATCATATATATACAAATCTTCAAAATACCTATCCCTAATA 1625
QY 493 TAGCAATTTTAAATAATTTTAAATAAATCATCTAAGCATTCAGATCTCTCA 552
DB 1624 AAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1565
QY 553 CACCTGTGATCTTATAGTCATGTATGTATGTAATCATCTAGTTGACACAGATTA 612
DB 1564 AATCATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1505
QY 613 AATTAATAGATTAATCTAGGAAATATATATATATATATATATATATATATATA 672
DB 1504 AATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1445
QY 673 ATCAATTAAGATTTTGTATTCCTCATGACATGACAA 708
DB 1444 ATCAAAAAAATCATCTTCAACGAAAAAATAAATA 1409

RESULT 4
US-10-204-708-20/c
Sequence 20, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIENSBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
PRIOR FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 20

LENGTH: 6866
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-20

Query Match 6.3%; Score 71.8; DB 4; Length 6866;
Best Local Similarity 48.0%; Pred. No. 6.2e-07;
Matches 205; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 108 AAAAAATTAATAAATAAATGCTTATCTTGTGTTTGTATTAATATATAGACTT 167
DB 1148 AAAAAAATTAATAAATCCCAAAAAATACAAATCCCTAATCAATTAATCAAAAAACT 1089
QY 168 AATCTTTATATATATATATAGTTTCTAGTCATAGACCACTCAGAGACAA 227
DB 1086 AAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1029
QY 228 GATTCAAGAAACATTTTGTAAACATCTTATAGAACTTTAGTTAGTTGAAG 287
DB 1028 AATTCGAAAAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 969
QY 288 TTAGATTAACAAAAAATTAACACGAGAAACATTAACCCATCAAGTCAAGT 347
DB 968 AATTTCTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 909
QY 348 ATCATAGAGTGAATGTTTGTATCTTAATTAATTAACACACACAAAAATACATTAAT 407
DB 908 CTAAATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATCT 909
QY 408 TATTAATATATATATATATATATTTTGTAAATCTTAGAGTTTCAAAACATCT 467
DB 848 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATC 789
QY 468 AATTCATGTTAGTTTATAGAAATATATATATTAATAAATAAATAAATAAATA 527
DB 788 AAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 729
QY 528 TTCTAAA 534
DB 728 ATAAAAA 722

RESULT 5
US-08-998-416-288/c
Sequence 288, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippson, Peter
APPLICANT: Pomlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
FILE REFERENCE: 1152
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416

APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
NUMBER OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 623926artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: Pf/S-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692RP
US-08-998-416-1137

Query Match 5.8%; Score 67; DB 3; Length 636;
Best Local Similarity 44.9%; Pred. No. 3.5e-06;
Matches 253; Conservative 0; Mismatches 310; Indels 0; Gaps 0;

QY 130 TTATCTCTTTCTGTTGTAATAATAATAGAGACTTAACTTTTAAATAATAATGTT 189
DB 636 TTTTATAAGTATTTTAACTACATCTTTATAATATTTTAAATAATAATGAT 577

QY 190 AATTAGGTTTCTAGTCTAGGACCACTGAGACAGATTTCAAGAAACAATTTGT 249
DB 576 AAAATATTAATTAATAATTAATAATAATAATAATAATAATAATAATAAT 517

QY 250 TAAACATCTTATTAGAACTTTTGTAGTCTTGAAGTTTAAAGTAAACAAAAAAT 309
DB 516 TATTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 457

QY 310 ACACGAGAAACAATAAACCCATACCTAGGTTTATCATAGAGTAATGTTTG 369
DB 456 TTTCTTAAAAAGATTTAAATTAATAATAATAATAATAATAATAATAATAATA 397

QY 370 ATATCATTAATAATAACACACAAAAATACATCTAATTAATAATAATAATAATA 429
DB 396 ATAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 337

QY 430 ATATTTTGTAAAAACTTAGAGTTTTCAAAAACAATCTTAATACATGATAGGTTTATAG 489

DB 336 ATTTTAAATAACAATAATAATAATAATAATAATAATAATAATAATAATAATA 277
QY 490 AAATACAAATATTTAAAAAATAATAATAATAATAATAATAATAATAATAATA 549
DB 276 TAAAGAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 217
QY 550 TCACACCTGCGTATCATTTAGTCAATGATGATGATGATGATGATGATGATG 609
DB 216 TAAATAGTATTCATATAATAATAATAATAATAATAATAATAATAATAATA 157
QY 610 TAAATAATAATAAGGATAACTAGGGAATATAATAATAATAATAATAATAATA 669
DB 156 TAAAGTAAATATATAATAATAATAATAATAATAATAATAATAATAATAATA 97
QY 670 AAATCAAAATTAGAATTTTGTAT 692
DB 96 ATAATGATAATAATAATAGTTTAAAT 74

RESULT 8
US-10-204-708-32/c
Sequence 32, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 32
LENGTH: 8093
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-32

Query Match 5.8%; Score 66.6; DB 4; Length 8093;
Best Local Similarity 46.2%; Pred. No. 9.5e-06;
Matches 222; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY 197 TTTTCTAGTCTAGCAGCACTCAGAGACAGATTTCAAGAAACAATTTTGTAAACAT 256
DB 2357 TTTTAAATCATCTTAAATTTTAACTTATACAAATAATAATAATAATAATA 2298

QY 257 CTTATTAGAACTTTTAGTTCAGTCTTGAAGTTAGAAATTAACAAAAAATAATACAC 316
DB 2297 AATTAAACATATAAAAAAATAATAATAATAATAATAATAATAATAATAATA 2238

QY 317 AGAAACATATAAACCCACTACCGTCTAGGTTATCATAGGATCAATGTTTTCATATCAT 376
DB 2237 CAATAATACCTTAAAGCTTATATCAATTAATTTTATTAATCAATTTTAAATACGTC 2178

QY 377 TAAATATAACACACAAAAATACATCTAATTAATAACAATATATGTTTATACATATTTT 436
DB 2177 GTACGTTAATAAAAAAATAATACTTCAATTTAAATAATAATAATAATAATAATA 2118

QY 437 TGTAAAACTTAGAGTTTTCAAAAACAATCTTAATACATGATAGGTTTATAGAAATACA 496


```

Db      2117 TATATAAATAATAACTATTTCCTCAAAACAATAAATACTAAAAAAAACAAATAAATAAAAAAACCA 2050
QY      497 AATATTTTAAAAAATAAATAATTTTAAAAAAAACATTTCTAAAGTCAITTCAGATCCTCTCACACC 556
Db      2057 CTAAAAAATAAAAAAAAAAATTAATAATAATAAATAAAAAAAAAAAAAACAATAAATAAT 1998
QY      557 TGTTGTCATCATTTAGTCATGTGTAGTACTACAAATCATTTGTAGTTTCACACAGAGTGTAATA 616
Db      1997 TTTTTCCTCTTTCACTTTATATAATAATAAACAATAAATAAATAAATAAATAAATACTTTATTTTCT 1938
QY      617 AATAAGGATATAAAGTAGGGATATATAATAATATATACAAATTAATAAATAAATAAAGGGAATAACA 676
Db      1937 ACTATATTAAATAAATAATTAACAATAAACAATAAACCATTATTATATATACCAATAACT 1878
QY      677 A 677
Db      1877 A 1877

RESULT 9
US-10-204-708-2/c
; Sequence 2, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BRELIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 10467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-2

Query Match          5.7%; Score 65.8; DB 4; Length 10467;
Best Local Similarity 45.2%; Pred. No. 1.6e-05;
Matches 284; Conservative 0; Mismatches 342; Indels 3; Gaps 1;

QY      91 AACGAAGGTGAAGAAGAAAAAGATTAAAAAAAATGTCCTTAICTCTTTGTTCTGTAAT 150
Db      1825 AAAAAAAAAAAAAAAAACAATAAACCCTTTAAATATAAAAAAAAAAAAAAACTAACCATAAATCCAAAT 1766
QY      151 AATAATATAAGAGACTTAAACCTTTTAAATAATAATTGTAATTAGGTTTTCTAGTCAATGA 210
Db      1765 AAATAAATAAATAATATATAATCAATCAATAAATAAATAAATAAATACTTACTTAACAACCT 1706
QY      211 GCACCACTCGAGACAAGNTTTCAGAAAAACAATTTTGTAAACATCTTATTAGAAACCTT 270
Db      1705 CACCGAAATCATACGMAAAAAACAATTCACCAAATAAAAAATAAATAATTAATAAACACCA 1646
QY      271 TTAGTTAAGTCTTCAGCTTGAATTAAACAAAAAAAEETACACACGAGAACACAAATAAA 330
Db      1645 AAAAATCAACATCAAAAAAAAACAAAAATCAAAAAAACTATACCCCAAAATACAAACTA 1586
QY      331 CCCCACTACCGGTGTATCATAGGATGAATGTTTTGTATATCATTTAAATAATAACACAC 390

```

```

1585 CCTATTTTATATAATAAAAATTTCATATAATAAATAAATTATATAAATAAAAATTTTACATATTAT 1522
391 ACAAATAATACATCTCAATATATAACAATATATGTGTATACATATATTTTTGTAAAAAATTAGA 450
1525 ATAAAAATCTATTTTATATCTACCAATACAAAACTATCTTAACCCCTTTATATAAAAAAAATTTAT 1466
451 GTTTTTCAAAACATCTCAATACATGATTTAGAGTTTATAGAAATACAAATATTTTAAAAAAT 510
1465 CAAACCACTAAAAAAAACCTTTACAAAATTTTAAATCTTTAAATATATAACCTTAATTAAATC 1406
511 ATAATTTTAAAAAAACATTTCTTAAAGTCATTTCAGATCTCTCACACCTGTGTGATCAATTTA 570
1405 ATAATAAAATTAATAAAAAATAAATACTAAACAAACCAATATATATAACA-----AAAAATAAAAAATA 1349
571 GTCATGTATGTAGTACAAATCATTGTAGTTTCACACGAGAGTAAATTAATAAGGATAAACT 630
1348 AAATCACAATAATTAATAAAATCCTTATACACAAACACCAATAAAAAATATATTACAAAAATA 1289
631 AGGGAATATATATAATATATACAAATTAATAAAAAAGGGAAAAATCAAAATAGAAATTTTGTG 690
1288 CTAAATATCCCTAAATATATCAAAATATAAAAATATAAAAAATAAAAAATCTAAATTTATAAAAAAA 1229
691 ATTCCCAATGACACAACTCCACATGCA 719
1228 AAATCCCACTCTACCTATACAACTCCA 1200

RESULT 10
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEIN
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 520 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Neda
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

US-08-487-826B-13

Query Match 5.7%; Score 65.8; DB 2; Length 19124;
Best Local Similarity 44.8%; Pred. No. 1.9e-05;
Matches 291; Conservative 0; Mismatches 357; Indels 1; Gaps 1;
QY 50 AACAAACAAATTTATGCTCTTCTGTTCTGTAATAATAATAGAGACTTTAAACCTTTAATA 119
DB 18279 AATTAAAGGAATAAAGCTTAATAATAATAATAAGAAATAGTATTTATTTATTAATAATAACA 18220
QY 120 AAAAATGTCCTTATCTCTTCTGTTCTGTAATAATAATAGAGACTTTAAACCTTTAATA 179
DB 18219 AGAAATATTATGTTATATTAATAATAATTTATTTAATAGGAACTATATATATGTTATT 18160
QY 180 TAAATATGTAATTAAGTTTCTAGTCTAGTACGACCACTCAGACAGAGATTTCAAGAA 239
DB 18159 ATAATAATTTTATATGAGATTTATATTTTCTCGTCGATTTATTCAGAAATTAGA 18100
QY 240 ACAATTTTGTATAACATCTTATTAGAACTTTTGTAGTTTAAAGTCTTGAAGT-TAGAAATTAAA 298
DB 18099 GTACAAAATAAACAATATATAAACAATATATAAATACACATTTTAAATATATAT 18040
QY 299 CAAAAAAATACACAGAGAAACAAATAAACCACTACCGTCAGGTTATCATTAAGAT 358
DB 18039 TATAAATAATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 17980
QY 359 GAAATGTTTGTATATCATTTAATAATAACACACACAAAATAACATCTTAATTAATAACAATAT 418
DB 17979 TGAAGGATATATATTTAGAAAGAGAAATATATATATATATATATATATATATATACG 17920
QY 419 AAGTTATACATATATTTTGTGAAAACTTAGAGTTTTCAAAACATCTCAATACATGAT 478
DB 17919 AAAAAATGTACAGAAAGAAACAAATAAACATAATATACATAATTTATTAATCAATAT 17860
QY 479 AGAGTTTATAGAAATACAAATATTTAAATAATATATTTTAAATAACATCTTAAGTCA 538
DB 17859 AGATACAAATTTATATCTAAATAATAAATAACATATATATATATATATATATATATAT 17800
QY 539 TTCAGATCTCTCACACCTGTGATCATTTTAGTCATGTATGTAGTACAAATCATTTGATG 598
DB 17799 ATTACAAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 17740
QY 599 TCACACAGAGTAAATAAATAAAGTAACATAGGATATATATATATATATATATATATATAT 658
DB 17739 TGATTATAACAAACAAACATATTTTTCACAAATATACCAAAACAAATAAATAAATAA 17680
QY 659 ATAAAAAGGGAATCAATCAATAGAAATTTTTCATTTCCCAATACACACA 707
DB 17679 ATAAATAAATAAAGCAATATACAAATATACAAATATATATATATATATATATATAT 17631

RESULT 11

US-10-204-708-50/c
; Sequence 50, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 50
; LENGTH: 6306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-50

Query Match 5.6%; Score 64.9; DB 4; Length 6306;
Best Local Similarity 46.3%; Pred. No. 2.2e-05;
Matches 315; Conservative 0; Mismatches 362; Indels 3; Gaps 3;
QY 10 TCATAGAAATGTTGTTTCTCTATCCACGACAAAGGGCAACAGTTTAAACAAACAA 69
DB 3409 TTAATAATAATCTTCTTAACCTACAAATAAACAAGCAITTAACATCTTAATCTCACA 3350
QY 70 ATTATATGTTTCATTTGAGATTTAAGGAAGTTAAGCAAGAAAAAGATTTAAAAAATGTC 129
DB 3349 ATACTCTTTAAAAAATAAATAAATAAATCACTATTACAAAAATCGCAACCAATCTCAC 3290
QY 130 TTATCTCTTTGTTTCTGTAATAATAATAATAAGAGACTTAACTTTTAAATATAATAATGT 189
DB 3289 AATTCCAATATCTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3230
QY 190 AATTAGGTTTCTAGTCATGAGCACTCTCAGAGACAAGATTTCAAGAAAAACAAATTTTGT 249
DB 3229 AATAAACCCTAATCAACATTAACGAAACCCCATCTTAACAAAAAATAAACAACAAACA 3170
QY 250 TAAACATCTTTATAGAACTTTTGTAGTTAAGTCTTGAAGTTAGAAATTA-AAACAAAAAAT 308
DB 3169 AAAAACTTTTATTAACCTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 3110
QY 309 TACACAGGAAACACAAATAAACCCCTACCGTCAGGTTATCATTAAGATGAAATGTTT 368
DB 3109 TAAATAAATAAATAAATCGGTTAACTCGAATTTAAATAAATAAATAAATAAATAAATAA 3050
QY 369 GATATCATTAATAATATACACACAAATAATACATCTTAATTTATAACAATA-TGTTATAC 427
DB 3049 ACTACACTCCAACTTAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2990
QY 428 ATATATTTTGTAAAAAATCTAGAGTTTTCABAACATCTTAATACATGATTAGAGTTTAT 487
DB 2989 ACATTAATATATATCTCTCCCTATATAATACATCAACTTACTTTTCAAAACAATAC 2930
QY 488 AGAAATCAAAATTTAAAAAATAATAATTTAAAAAATAAATAAATAAATAAATAAATAAATA 547
DB 2929 TAAATAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2870
QY 548 TCTCACACCTGTGATCATTTAGTCAATGTATGTAGTACAAATCAATGTTAGTTCACAACAG 607
DB 2869 AACCAACACTTTTATATACACCACTAACACATTTCTAC-ATCTATTATAAATAATCCTT 2811
QY 608 AGTAAAAATAAATAAGGATAAATAAGGATAATATATAATATATAATATAAATAAATAAATA 667
DB 2810 TAAATAACCGTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2751
QY 668 GGAATAATCAAAATAGAAATTT 687
DB 2750 AAAAACTATCAATTAATTTAT 2731

RESULT 12

US-10-204-708-12/c
; Sequence 12, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1


```
Query Match      5.6%; Score 63.8; DP 1; Length 660;
Best Local Similarity 48.5%; Pred. No. 1.8e-05;
Matches 272; Conservative 0; Mismatches 277; Indels 12; Gaps 3;

QY 109 AAAAGATTAAAAAATCCTCTATCTCTTTGTTCTGTAATAATAATAAGAGACTTA 168
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 637 ATAAATTTTAAATAATATAAATTTCTTTATGATTGATAATCAATAAATGAGTTA 578
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 169 AACTTTTAAATAATAATGTAATTAGGTTTCTAGTCATGAGCACCACCTCAGAGACAAG 228
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 577 TACITTTTGGAAATTTTATCAATGATATATTTTITTTAAACATTGAAGATATACTA 518
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 229 ATTCAGAGAAACAATTTTGTAAACATCTTATTAGAAACTTTTAGTTAGTCTTGAAGT 288
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 517 ATTTTAAATAATAATTTTCTATAAATTTTATAATCTTTTATTTT--GTAATTGATTC 460
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 TAGAATTAACAAAAAATTTACACAGAGAAACACAATAAACCCACTACCGTCAGGTTA 348
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 459 CATATTAACTCACTACTACTAATAGGAAACACATTATTAAAGTTACCAATTTATTTT 400
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 TCATAAGGATCAAAATGTTTTCATATCATTAATAATAACACACACAAAAATACATCTAAT 408
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 399 AGATAATTATAATTTTAAATTTACTAAATTTATTA-----ATAAAATTATAGTCATTT 345
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 409 ATACCAATATATGTTATACATATTTTGTAAAAAATTTAGAGTTTTCAAAAACATCTA 468
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 344 ATATTACATGATTCACAAATTTAAAAATTTCTATAGAAATGAGTGTAGTATATTTACTATA 285
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 ATACATGATTAGAGTTTATAGAAATACAAATATTATAAAAAATATAATTTTAAAAAACA 528
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 284 TTGCTATTTTGTATATAAGATATCTATAATATGTTATTTTAAATTTTGTATATAAA 225
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 TCTAAAGTCATTCAGATCCTCTCACACCTGTGTATCATTTTAGTCATGTATAGTA--- 585
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 224 TTAAATTAATAATTTTAAATTTTGAATAATAATAAACTTTTAATATTTCTGGAAATATA 165
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 586 --CAATCATTCGTAGTTCAACAGAGTAAATAATAAGGATAAACTAGGGAATATATAT 643
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 TTAAATAATTTATTCATATATATATATATGCAATTCCTCTAAATTAACATAATTTT 105
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 644 AATATATCAATTAATAAAA 664
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 104 AATATATTAATTAATAACA 84
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: June 2, 2004, 02:26:17
Job time : 112 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 01:00:19 ; Search time 538 Seconds
(without alignments)
9064.928 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148
Sequence: 1 taggaccccttcataagaaaa.....agagagtgatggttaatgca 1148

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_29Jan04:*
2: Geneseqn1980s:*
3: Geneseqn1990s:*
4: Geneseqn2000s:*
5: Geneseqn2001as:*
6: Geneseqn2002bs:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	1148	6	ABN83916 Truncated
2	1148	100.0	1832	6	ABN83929 Arcelin-5
3	1144.8	99.7	1821	6	ABD39066 Phaseolus
4	1122	97.7	1122	6	ABN83917 Truncated
5	796.6	69.4	1866	6	ABN83928 Arcelin-4
6	795	69.3	1872	6	ABN83927 Arcelin-3
7	90	7.8	8056	7	ABZ10246 Haematopo
8	87	7.6	3693	7	ABZ10199 Haematopo
9	85.2	7.4	4846	6	ABK87142 Scarlet r
10	85	7.4	2888	6	ABN83926 Arcelin-5
11	85	7.4	8056	7	ABZ10100 Haematopo
12	84.8	7.4	8056	7	ABZ10246 Haematopo
13	84.4	7.3	1211	2	AAQ94051 Kidney be
14	83.4	7.3	5925	6	ABL33576 Human inm
15	83	7.2	26997	4	AA546748 Tumour su
16	82.6	7.2	1501	7	ABZ10188 Haematopo
17	82.6	7.2	1501	9	AD84162 Human lym
18	82.2	7.2	16786	6	ABL34157 Human inm
19	81.4	7.1	3683	7	ABZ10053 Haematopo
20	81.2	7.1	6216	6	ABK39932 Human che
21	81.2	7.1	6216	6	ABL70139 Chemical
22	80.6	7.0	14006	6	ABL33958 Human inm
23	80.2	7.0	6048	6	ABQ67002 Human ang

ALIGNMENTS

RESULT 1

ABN83916

ID ABN83916 standard; DNA; 1148 BP.

XX AC

XX ABN83916;

XX DT 06-SEP-2002 (first entry)

XX DE Truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.

XX KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;

XX KM pharmaceutical; ds.

XX OS Phaseolus vulgaris.

XX XX WO200250295-A2.

XX PD 27-JUN-2002.

XX PP 17-DEC-2001; 2001WO-US047495.

XX PR 18-DEC-2000; 2000US-0255879P.

XX (RENE-) RENESSEN LLC.

XX Wang Q, Dubois P, Liang J, Oulmassov T;

XX WPI; 2002-508809/54.

XX New transformed or transgenic soybeans plants or cells with an Arcelin-5

XX promoter, for use as an improved dietary source of protein for humans or

XX animals, or for producing soybeans with important agricultural or

XX nutritional properties.

XX Claim 4; Page 65-66; 74pp; English.

XX The invention relates to a transformed soybean plant cell and transgenic

XX soybean plant, both of which has a nucleic acid molecule comprising the

XX phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.

XX The transformed soybean plant cell and transgenic soybean plant are

XX useful as an improved source of dietary protein for humans and livestock.

XX These are also useful for producing soybean plants that exhibit important

XX agricultural, nutritional or pharmaceutical properties. The current

XX sequence represents the truncated P. vulgaris exotic genotype G02771

XX Arcelin-5 promoter sequence. This sequence is a deletion mutant of the

XX Arcelin-5 promoter, created by the removal of approximately 600 base

C 24 80 7.0 83391 6 ABQ67093 Abq67093 Human ang

C 25 79.8 7.0 8056 7 ABZ10100 Abz10100 Haematopo

C 26 79.8 7.0 21354 4 AAS46815 Aas46815 Tumour su

C 27 79.2 6.9 6644 2 AAX33181 Aax33181 Base sequ

C 28 79.2 6.9 6568 6 ABL33397 Abl33397 Human inm

C 29 79.2 6.9 7372 2 AAX33182 Aax33182 Base sequ

C 30 79.2 6.9 7797 2 AAX33180 Aax33180 Cowpox vi

C 31 79.2 6.9 7892 6 ABK40056 Abk40056 Human che

C 32 79.2 6.9 7996 2 AAX33184 Aax33184 Base sequ

C 33 79.2 6.9 14615 4 AAS46704 Aas46704 Tumour su

C 34 79 6.9 6106 4 AAS46429 Aas46429 Tumour su

C 35 79 6.9 6106 6 ABK40031 Abk40031 Human che

C 36 79 6.9 6106 6 ABL33472 Abl33472 Human inm

C 37 78.6 6.8 778 6 ABQ15588 Abq15588 Oligonucle

C 38 78.6 6.8 778 6 ABQ15589 Abq15589 Oligonucle

C 39 78.6 6.8 12237 6 ABL34358 Abl34358 Human inm

C 40 78.4 6.8 8693 6 ABL32808 Abl32808 Human inm

C 41 78.2 6.8 5654 4 AAS46623 Aas46623 Tumour su

C 42 78.2 6.8 5854 6 ABL33874 Abl33874 Human inm

C 43 78.2 6.8 6175 6 ABL33307 Abl33307 Human inm

C 44 77.8 6.8 15548 6 ABL34155 Abl34155 Human inm

C 45 77.8 6.8 18218 6 ABL33949 Abl33949 Human inm

CC pairs from the promoter nucleic acid sequence
SQ Sequence 1148 BP; 420 A; 211 C; 155 G; 362 T; 0 U; 0 Other;

Query Match 100.0%; Score 1148; DB 6; Length 1148;
Best Local Similarity 100.0%; Pred. No. 4.5e-190;
Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGATCCTTCCTCAATAGAAAATGCTGTTATTTCTCATCCACAGACAAAGGGGCAACAGTTA 60
DB 1 TAGGATCCTTCCTCAATAGAAAATGCTGTTATTTCTCATCCACAGACAAAGGGGCAACAGTTA 60

QY 61 ACAAAAACAAATTTATGTTTCTATTTTCAATTAAGGAAGTAAAGGAAGAAAAGATTAAAA 120
DB 61 ACAAAAACAAATTTATGTTTCTATTTTCAATTAAGGAAGTAAAGGAAGAAAAGATTAAAA 120

QY 121 AAAATGTCCTTATCTCTTTGTTTCTGTAATTAATTAATTAAGAGACTTAAACTTTTAAATAT 180
DB 121 AAAATGTCCTTATCTCTTTGTTTCTGTAATTAATTAATTAAGAGACTTAAACTTTTAAATAT 180

QY 181 AATAATGTAATTAGGTTTCTTAGTCATGAGCACCACCTCAGAGACAAGATTTCAGAAAA 240
DB 181 AATAATGTAATTAGGTTTCTTAGTCATGAGCACCACCTCAGAGACAAGATTTCAGAAAA 240

QY 241 CAATTTGTTTAAACATCTTATAGAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAAACA 300
DB 241 CAATTTGTTTAAACATCTTATAGAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAAACA 300

QY 301 AAAAAAATTTACACGAGAAACACATAAAACCACTACCGTCAGGTTATCATAGGATGA 360
DB 301 AAAAAAATTTACACGAGAAACACATAAAACCACTACCGTCAGGTTATCATAGGATGA 360

QY 361 AATGTTTGTATATCAATTAATATAACACACACAAATAATACATCTAATTAATTAACAATAT 420
DB 361 AATGTTTGTATATCAATTAATATAACACACACAAATAATACATCTAATTAATTAACAATAT 420

QY 421 GTTATACATATATTTTCTGTAAGCTTAGCTTTTCAAAACATCTTAATACATGATTAG 480
DB 421 GTTATACATATATTTTCTGTAAGCTTAGCTTTTCAAAACATCTTAATACATGATTAG 480

QY 481 AGTTTATAGAAATCAAAATATTTAAATAATATAATTTTAAATAAAACATTTCTAAAGTCATT 540
DB 481 AGTTTATAGAAATCAAAATATTTAAATAATATAATTTTAAATAAAACATTTCTAAAGTCATT 540

QY 541 CAGATCCTCTCACACCTGTGTGATCAATTTAGTCATGATGTAGTACATCAATCTAGTTTC 600
DB 541 CAGATCCTCTCACACCTGTGTGATCAATTTAGTCATGATGTAGTACATCAATCTAGTTTC 600

QY 601 ACAACAGAGTAAATAAATAGGATAAACTAGGGAATATATATATATATACAAATTAAAT 660
DB 601 ACAACAGAGTAAATAAATAGGATAAACTAGGGAATATATATATATATATACAAATTAAAT 660

QY 661 AAAAAAGGAAATCAAAATTTAGAAATTTTGTATTTCCCAATGACACAACTCACCAATGAC 720
DB 661 AAAAAAGGAAATCAAAATTTAGAAATTTTGTATTTCCCAATGACACAACTCACCAATGAC 720

QY 721 GCTGCGACCTGAGCTCCCTCTCCACACATGCTCATGTCATCTTTCGATCTTGGCTTT 780
DB 721 GCTGCGACCTGAGCTCCCTCTCCACACATGCTCATGTCATCTTTCGATCTTGGCTTT 780

QY 781 TTCATATGACACAACTCGCCATGATGTTGCGACGTGAGCTCCTTCTCTTCCCATGAT 840
DB 781 TTCATATGACACAACTCGCCATGATGTTGCGACGTGAGCTCCTTCTCTTCCCATGAT 840

QY 841 GACACACCTGGCATGATGTCGCCACCTCAGCTCCGACCTCTTCTCATTTAGCCCTAC 900
DB 841 GACACACCTGGCATGATGTCGCCACCTCAGCTCCGACCTCTTCTCATTTAGCCCTAC 900

QY 901 TGGCCATGACACTGCCACCTCAGCACTCCCTCTCACTTCCCATCTGCTACCTGCCAAACCG 960
DB 901 TGGCCATGACACTGCCACCTCAGCACTCCCTCTCACTTCCCATCTGCTACCTGCCAAACCG 960

QY 961 CTTCTCTCCATAAATATCTATTTAAATTTAACTAAATTTTTCATATATCTTTTGTATGA 1020

DB 961 CTTCTCTCCATAAATATCTATTTAAATTTAACTAAATTTTTCATATATCTTTTGTATGA 1020

QY 1021 CGTGGATGCAATGCCATCGTGTGTTTAAATTAATTTGTTTAAATTTGAGTTGAATAATAATGA 1080

DB 1021 CGTGGATGCAATGCCATCGTGTGTTTAAATTAATTTGTTTAAATTTGAGTTGAATAATAATGA 1080

QY 1081 AAGAAAAAGTTTGGAAAGATTTTTCGATTTGTTCTTCTATAAATAGAGAGAGAGTGTATGG 1140

DB 1081 AAGAAAAAGTTTGGAAAGATTTTTCGATTTGTTCTTATAAATAGAGAGAGAGTGTATGG 1140

QY 1141 TTAATGCA 1148
DB 1141 TTAATGCA 1148

RESULT 2
ABN83929
ID ABN83929 standard; DNA; 1832 BP.
XX AC ABN83929;
XX AC
XX 06-SEP-2002 (first entry)
XX DE Arcelin-5 full length promoter sequence.
XX DE Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;
XX KW pharmaceutical; ds.
XX OS Phaseolus vulgaris.
XX FN WO200250295-A2.
XX 27-JUN-2002.
XX 17-DEC-2001; 2001WO-US047495.
XX 18-DEC-2000; 2000US-0255879P.
XX (RENE-) RENESSEN LLC.
XX Wang Q, Dubois P, Liang J, Oulmassov T;
XX WPI; 2002-508809/54.
XX
XX New transformed or transgenic soybeans plants or cells with an Arcelin-5
XX promoter, for use as an improved dietary source of protein for humans or
XX animals, or for producing soybeans with important agricultural or
XX nutritional properties.
XX Example 1; Fig 4; 74pp; English.
XX
XX The invention relates to a transformed soybean plant cell and transgenic
XX soybean plant, both of which has a nucleic acid molecule comprising the
XX Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
XX The transformed soybean plant cell and transgenic soybean plant are
XX useful as an improved source of dietary protein for humans and livestock.
XX These are also useful for producing soybean plants that exhibit important
XX agricultural, nutritional or pharmaceutical properties. The current
XX sequence represents an arcelin-5 full length promoter sequence

XX SQ Sequence 1832 BP; 700 A; 311 C; 214 G; 607 T; 0 U; 0 Other;

Query Match 100.0%; Score 1148; DB 6; Length 1832;
Best Local Similarity 100.0%; Pred. No. 4.5e-190;
Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGATCCTTCCTCAATAGAAAATGTTATTTCTCATCCACAGACAAAGGGGCAACAGTTA 60
DB 659 TAGGATCCTTCCTCAATAGAAAATGTTATTTCTCATCCACAGACAAAGGGGCAACAGTTA 718

QY 61 ACAAAAACAAATTTATGTTTCTATTTTCAATTAAGGAAGTAAAGGAAGAAAAGATTAAAA 120

Db 719 ACAAAACAATTTATGTTTCATTTGAGATTAGGAGGTAAGGAGAGAAAAAGATTAAAA 778
Qy 121 AAAATGCTCTTATCTCTTTCTGTTGTAATAATAATAAAGAGACCTTAACCTTTTATAT 180
Db 779 AAAATGCTCTTATCTCTTTCTGTTGTAATAATAATAAAGAGACCTTAACCTTTTATAT 838
Qy 181 AATAATTTGTAATAGGTTTCTAGTCATGAGCACCACCTCAGAGACAGAGATTTCAAGAAAA 240
Db 839 AATAATTTGTAATAGGTTTCTAGTCATGAGCACCACCTCAGAGACAGAGATTTCAAGAAAA 898
Qy 241 CAATTTGTTAAACATCTTATAGAACTTTAGTTAGTCTTGAAGTTAGAAATTAACA 300
Db 899 CAATTTGTTAAACATCTTATAGAACTTTAGTTAGTCTTGAAGTTAGAAATTAACA 958
Qy 301 AAAAAAATTAACACAGAGAGAAACAATAAACCCTACCTACCGTTCAGTTATCATAGGATGA 360
Db 959 AAAAAAATTAACACAGAGAGAAACAATAAACCCTACCGTTCAGTTATCATAGGATGA 1018
Qy 361 AATGTTTGTATATCAATTAATAATAACACACACAAAAATACATCTTAATAAACAATAT 420
Db 1019 AATGTTTGTATATCAATTAATAATAACACACACAAAAATACATCTTAATAAACAATAT 1078
Qy 421 GTTATACATATATTTTGTAAACCTTAGAGTTTTCAAAACATCTTAATAACATGATTAG 480
Db 1079 GTTATACATATATTTTGTAAACCTTAGAGTTTTCAAAACATCTTAATAACATGATTAG 1138
Qy 481 AGTTTATAGAAATACAAATATTTTAAAAAATATATATTTTAAAAAACAATCTTAAGTCATT 540
Db 1139 AGTTTATAGAAATACAAATATTTTAAAAAATATATATTTTAAAAAACAATCTTAAGTCATT 1198
Qy 541 CAGATCCTCTCACACCTGTGTGATCATTTAGTCAATGATGTAGTACATCAATGTAGTTC 600
Db 1199 CAGATCCTCTCACACCTGTGTGATCATTTAGTCAATGATGTAGTACATCAATGTAGTTC 1258
Qy 601 ACAACAGATTAATAATAAGGATAAATAGGATAAATAGGATAATATATATATATACAAATTAAT 660
Db 1259 ACAACAGATTAATAATAAGGATAAATAGGATAAATAGGATAATATATATATATACAAATTAAT 1318
Qy 661 AAAAAAGGAAAAATCAATATTAGAAATTTTGAATCCCAACATGACACAACTCACATGCAC 720
Db 1319 AAAAAAGGAAAAATCAATATTAGAAATTTTGAATCCCAACATGACACAACTCACATGCAC 1378
Qy 721 GCTGCACTCTCAGTCCCTCTCTCCACATGCTCATGTCATCTTTGCGACTTTGGCTTT 780
Db 1379 GCTGCACTCTCAGTCCCTCTCTCCACATGCTCATGTCATCTTTGCGACTTTGGCTTT 1438
Qy 781 TTCATATGACAACTCGCCATGCTGTTGCCAGTGCCTCTCTCTCTTCCCATGAT 840
Db 1439 TTCATATGACAACTCGCCATGCTGTTGCCAGTGCCTCTCTCTCTTCCCATGAT 1498
Qy 841 GACACCACTGGGATGCAATGCTGCCACTCAGCTCCCACTCTCTCTCATTTAGCCCTAC 900
Db 1499 GACACCACTGGGATGCAATGCTGCCACTCAGCTCCCACTCTCTCTCATTTAGCCCTAC 1558
Qy 901 TGGCCATGCACTGCCACCTCAGCACTCTCTCTCACTTCCCATTTGCTACCTGCCAAACCG 960
Db 1559 TGGCCATGCACTGCCACCTCAGCACTCTCTCTCACTTCCCATTTGCTACCTGCCAAACCG 1618
Qy 961 CTTCTCTCCATTAATATCTATTAAATTTAAATTTAACTTAATTTTCAATATCTTTTGTATGA 1020
Db 1619 CTTCTCTCCATTAATATCTATTAAATTTAAATTTAACTTAATTTTCAATATCTTTTGTATGA 1678
Qy 1021 CGTGGATGCACTGGATGTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTAAATTTAA 1080
Db 1679 CGTGGATGCACTGGATGTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 1738
Qy 1081 AAAAAAAGTTGGAAGATTTGCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
Db 1739 AAAAAAAGTTGGAAGATTTGCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1798
Qy 1141 TTAATGCA 1148
Db 1799 TTAATGCA 1806

RESULT 3
AAD29066
ID AAD29066 standard; DNA; 1821 BP.
XX
AC AAD29066;
XX
DT 07-MAY-2002 (first entry)
XX
DE Phaseolus vulgaris arcelin promoter.
XX
KW Heterologous gene expression; plant; arcelin promoter; arcelin leader;
KW seed preferred expression cassette; ds.
XX
OS Phaseolus vulgaris.
XX
PN W0200200899-A2.
XX
PD 03-JAN-2002.
XX
PF 31-MAY-2001; 2001WO-EP006298.
XX
PR 29-JUN-2000; 2000EP-00202278.
XX
PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOL.
XX
PI Angenon G, De Jaeger G, Goossens A, Depicker A;
XX
DR WPI; 2002-139925/18.
XX
PT Novel seed preferred expression cassette having gene regulatory elements
PT comprising arcelin promoter, arcelin 5' leader, and arcelin 5' 3' end,
PT useful for heterologous gene expression in plants.
XX
PS Claim 1; Page 46-47; 52pp; English.
XX
CC The invention relates to heterologous gene expression in plants. The
CC invention also relates to seed preferred expression cassette having gene
CC regulatory elements comprising arcelin promoter, arcelin 5' leader, and
CC arcelin 5' 3' end. This expression cassette is useful for heterologous
CC gene expression in plants. The protein encoded by the heterologous gene
CC is a single chain antibody variable fragment (scfv). The present sequence
CC is Phaseolus vulgaris arcelin promoter of the invention
XX
SQ Sequence 1821 BP; 598 A; 308 C; 210 G; 605 T; 0 U; 0 Other;
Query Match 99.7%; Score 1144.8; DB 6; Length 1821;
Best Local Similarity 99.8%; Pred. No. 1.7e-189;
Matches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TAGGATCCTTCAATAGAAATGTTTATTTCTCATCCAGACAAAAGGGCCACAGTTA 60
Db 674 TAGGATCCTTCAATAGAAATGTTTATTTCTCATCCAGACAAAAGGGCCACAGTTA 733
Qy 61 AAAAAACAAATTTATGTTTCATTTGAGATTAGGAGGTAAGGAGAGAAAAAGATTAAAA 120
Db 734 AAAAAACAAATTTATGTTTTCATTTGAGATTAGGAGGTAAGGAGAGAAAAAGATTAAAA 793
Qy 121 AAAATGCTCTTATCTCTTTCTGTTGTAATAATAATAAAGAGACCTTAACCTTTTATAT 180
Db 794 AAAATGCTCTTATCTCTTTCTGTTGTAATAATAATAAAGAGACCTTAACCTTTTATAT 853
Qy 181 AATAATTTGTAATAGGTTTCTAGTCATGAGCACCCTCAGAGACAGAGATTTCAAGAAAA 240
Db 854 AATAATTTGTAATAGGTTTCTAGTCATGAGCACCCTCAGAGACAGAGATTTCAAGAAAA 913
Qy 241 CAATTTGTTAAACATCTTATAGAACTTTTAGTTAGTCTTGAAGTTAGAAATTAACA 300
Db 914 CAATTTGTTAAACATCTTATAGAACTTTTAGTTAGTCTTGAAGTTAGAAATTAACA 973
Qy 301 AAAAAAATTAACACAGAGAGAAACAATAAACCCTACCTCAGGTTTATCATAGGATGA 360

DB 974 AAAAAAGTACACGAGGAAACACAATAAACCCACTACCGTCAGGTTATCATAGAGTGA 1033
QY 361 AATGTTTGTATATCATTAATAAATACACACACAAAAATACATCTAATTAATAACAATATAT 420
DB 1034 AATGTTTGTATATCATTAATAAATACACACACAAAAATACATCTAATTAATAACAATATAT 1093
QY 421 GTTATACATATATTTTGTAAATACTAGAGTTTTCAAAACATCTCTAATACATGATTAG 480
DB 1094 GTTATACATATATTTTGTAAATACTAGAGTTTTCAAAACATCTCTAATACATGATTAG 1153
QY 481 AGTTTATAGATACAAATATTTTAAATAATAAATTTTAAATAAATCTTAAGTCAAT 540
DB 1154 AGTTTATAGATACAAATATTTTAAATAATAAATTTTAAATAAATCTTAAGTCAAT 1213
QY 541 CAGATCTCTCACACTGTGTGATCATTTAGTCATGTATGTAGTCAATCAATTTAGTTC 600
DB 1214 CAGATCTCTCACACTGTGTGATCATTTAGTCATGTATGTAGTCAATCAATTTAGTTC 1273
QY 601 ACAACAGATGAATAAATAAAGGATAAAGGATAAAGGATAAATAAATAAATAAATAAATAA 660
DB 1274 ACAACAGATGAATAAATAAAGGATAAAGGATAAAGGATAAATAAATAAATAAATAAATAA 1333
QY 661 AAAAAAGGAAATCAAAATAGAAATTTTGTATCCCAATGACACAACTCAACCATGAC 720
DB 1334 AAAAAAGGAAATCAAAATAGAAATTTTGTATCCCAATGACACAACTCAACCATGAC 1393
QY 721 GTTGACATGATCT 780
DB 1394 GTTGACATGATCT 1453
QY 781 TTTCACTATGACACAACTGCCATGATGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGT 840
DB 1454 TTTCACTATGACACAACTGCCATGATGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGT 1513
QY 841 GACACATGTGGCATGATGCTGCGACCTGACCTGCGACCTGCTCTCTCTCTCTCTCTCTCTCT 900
DB 1514 GACACATGTGGCATGATGCTGCGACCTGACCTGCGACCTGCTCTCTCTCTCTCTCTCTCTCT 1573
QY 901 TGGCATGACACACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 960
DB 1574 TGGCATGACACACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 1633
QY 961 CTCTCTCTCAATAATATCTATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 1020
DB 1634 CTCTCTCTCAATAATATCTATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 1693
QY 1021 CTTGGATGATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 1694 CTTGGATGATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1753
QY 1081 AAGAAAAAGTTGGAAGATTTTGCATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 1140
DB 1754 AAGAAAAAGTTGGAAGATTTTGCATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 1813
QY 1141 TTAATGCA 1148
DB 1814 TTAATGCA 1821

RESULT 4
ABN83917
ID ABN83917 standard; DNA; 1122 BP.
XX AC ABN83917;
XX AC ABN83917;
DT 06-SEP-2002 (first entry)
XX Truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter #2.
DE Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;
XX pharmaceutical; ds.
OS Phaseolus vulgaris.

XX WO200250295-A2.
FN 27-JUN-2002.
XX 17-DEC-2001; 2001WO-US047495.
XX 18-DEC-2000; 2000US-0255879P.
PR (RENE-) RENESSEN LLC.
XX Wang Q, Dubois P, Liang J, Oulmassov T;
PI WPI; 2002-508809/54.
DR New transformed or transgenic soybeans plants or cells with an Arcelin-5
PT promoter, for use as an improved dietary source of protein for humans or
PT animals, or for producing soybeans with important agricultural or
PT nutritional properties.
XX Example 5; Page 66; 74pp; English.
XX The invention relates to a transformed soybean plant cell and transgenic
CC soybean plant, both of which has a nucleic acid molecule comprising the
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
CC The transformed soybean plant cell and transgenic soybean plant are
CC useful as an improved source of dietary protein for humans and livestock.
CC These are also useful for producing soybean plants that exhibit important
CC agricultural, nutritional or pharmaceutical properties. The current
CC sequence represents a truncated P. vulgaris exotic genotype G02771
CC Arcelin-5 promoter sequence. This sequence was used in the expression of
CC a GUS reporter gene in a soybean cotyledon
XX SQ Sequence 1122 BP; 410 A; 210 C; 146 G; 356 T; 0 U; 0 Other;

Query Match 97.7%; Score 1122; DB 6; Length 1122;
Best Local Similarity 100.0%; Pred. NO. 1.4e-185; Indels 0; Gaps 0;
Matches 1122; Conservative 0; Mismatches 0;
QY 3 GGATCCTTCAATAGAAATGTTTATTTTCTCATCACCAGACAAAGGGGCAACGTTTAAAC 62
DB 1 GGATCCTTCAATAGAAATGTTTATTTTCTCATCACCAGACAAAGGGGCAACGTTTAAAC 60
QY 63 AAAAAAATTTATGTTTTCATTTGAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 122
DB 61 AAAAAAATTTATGTTTTCATTTGAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 120
QY 123 AATGTCCTTATCTCTTGTGTTTCTGTAATAATAATAATAATAATAATAATAATAATAATAATA 182
DB 121 AATGTCCTTATCTCTTGTGTTTCTGTAATAATAATAATAATAATAATAATAATAATAATAATA 180
QY 183 TAATGTAAATTAAGTGTGTTTCTAGTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 242
DB 181 TAATGTAAATTAAGTGTGTTTCTAGTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
QY 243 ATTTTGTGTAACATCTTATTAGAAATCTTTAGTTTAAAGTCTTGAAGTTAGAAATTAACAAA 302
DB 241 ATTTTGTGTAACATCTTATTAGAAATCTTTAGTTTAAAGTCTTGAAGTTAGAAATTAACAAA 300
QY 303 AAAAAATTACACACGAGAAACAAATAAACCCACTACCGTCAGGTTATCATAGGATGA 362
DB 301 AAAAAATTACACACGAGAAACAAATAAACCCACTACCGTCAGGTTATCATAGGATGA 360
QY 363 TGTGTTGATATCATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 422
DB 361 TGTGTTGATATCATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 420
QY 423 TATACATATATTTTGTAAATACTTAGAGTTTTCAAAACATTTCTAATAATGATTAG 482
DB 421 TATACATATATTTTGTAAATACTTAGAGTTTTCAAAACATTTCTAATAATGATTAG 480
QY 483 TTTATAGAAATACAAATATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 542

481	Db	TTTATAGAATAACAAATATTTTAAAAAATAATATTTTAAAAAACAATTTCTAAAGCTATTC	540
543	Qy	GATCCTCTCACACCTCTGTGATCATTTAGTATGATGTAGTACAAATCATTTGTAGTTCAC	602
541	Db		
541	Db	GATCCTCTCACACCTCTGTGATCATTTAGTATGATGTAGTACAAATCATTTGTAGTTCAC	600
603	Qy	AACAGAGTAAATAATATAGGATAAACTAGGGAATATATATATATATACAAATTAATAA	662
601	Db		
601	Db	AACAGAGTAAATAATATAGGATAAACTAGGGAATATATATATATATACAAATTAATAA	660
663	Qy	AAAAAGGAAAAATCAAATTTAGAAATTTTGATTTCCCAATGACACAACTCACCATGCAACGC	722
661	Db		
661	Db	AAAAAGGAAAAATCAAATTTAGAAATTTTGATTTCCCAATGACACAACTCACCATGCAACGC	720
723	Qy	TGCCACTCAGCTCCCTCTCTCCACACATGCTCTCATGTCACCTTTGCGACTTTGGCTTTTT	782
721	Db		
721	Db	TGCCACTCAGCTCCCTCTCTCCACACATGCTCTCATGTCACCTTTGCGACTTTGGCTTTTT	780
783	Qy	CACATGACACAACCTCGCCATGATGTTGCCAGTGGAGCTCCTTCCTTCCCATGATGA	842
781	Db		
781	Db	CACATGACACAACCTCGCCATGATGTTGCCAGTGGAGCTCCTTCCTTCCCATGATGA	840
843	Qy	CACCACCTGGGCATGCTATGTCGCCACCTCAGCTCCCACTCTTCTCATATATGAGCCTACTG	902
841	Db		
841	Db	CACCACCTGGGCATGCTATGTCGCCACCTCAGCTCCCACTCTTCTCATATATGAGCCTACTG	900
903	Qy	GCCATGACACATGCCACCTCAGCAGCTCCTCTCAGCTCCCACTGCTACTGCCAAAACCGCT	962
901	Db		
901	Db	GCCATGACACATGCCACCTCAGCAGCTCCTCTCAGCTCCCACTGCTACTGCCAAAACCGCT	960
963	Qy	TCCTCCATAAATATCTATTTAAATTTTAAACTAAATTTTATATATCTTTTGTATGACG	1022
961	Db		
961	Db	TCCTCCATAAATATCTATTTAAATTTTAAACTAAATTTTATATATCTTTTGTATGACG	1020
1023	Qy	TGATGCAATTGCCATCGTCTGTTTAATTAATTTGTTAAATTTGGAGTTGANTATATAAATGAAA	1082
1021	Db		
1021	Db	TGATGCAATTGCCATCGTCTGTTTAATTAATTTGTTAAATTTGGAGTTGANTATATAAATGAAA	1080
1083	Qy	GAATAAGTTGGAAGATTTTGCATTTTGTGTTGTTATATAATA	1124
1081	Db		
1081	Db	GAATAAGTTGGAAGATTTTGCATTTTGTGTTGTTATATAATA	1122
RESULT 5			
ABN83928			
ID	ABN83928	standard; DNA; 1866 BP.	
XX	XX		
AC	ABN83928;		
XX	XX		
DT	DT		
XX	XX		
DE	DE		
XX	XX	Arcelin-4 full length promoter sequence.	
KW	KW	Arcelin-4; promoter; plant; transgenic; soybean; agriculture; nutrition;	
KW	KW	pharmaceutical; ds.	
XX	XX		
OS	OS	Phaseolus vulgaris.	
XX	XX		
PN	PN	WO200250295-A2.	
XX	XX		
PD	PD	27-JUN-2002.	
XX	XX		
PF	PF	17-DEC-2001; 2001WO-US047495.	
XX	XX		
PR	PR	18-DEC-2000; 2000US-0255879P.	
XX	XX		
PA	PA	(RENE-) RENESSEN LLC.	
PI	PI	Wang Q, Dubois P, Liang J, Oulmassov T;	
XX	XX		
DR	DR	WPI; 2002-508809/54.	
XX	XX		
PT	PT	New transformed or transgenic soybeans plants or cells with an Arcelin-5	

promoter, for use as an improved dietary source of protein for humans or animals, or for producing soybeans with important agricultural or nutritional properties.

Example 1; Fig 4; 74pp; English.

The invention relates to a transformed soybean plant cell and transgenic soybean plant, both of which has a nucleic acid molecule comprising the Phaseolus vulgaris exotic genotype G02771 Arceiin-5 promoter sequence.

The transformed soybean plant cell and transgenic soybean plant are useful as an improved source of dietary protein for humans and livestock. These are also useful for producing soybean plants that exhibit important agricultural, nutritional or pharmaceutical properties. The current sequence represents an arceiin-4 full length promoter sequence

Sequence 1866 BP; 692 A; 339 C; 209 G; 626 T; 0 U; 0 Other;

Query Match	69.4%;	Score 796.6;	DB 6;	Length 1866;
Best Local Similarity	85.4%;	Pred. No. 3.9e-129;		
Matches 1034;	Conservative 0;	Mismatches 99;	Indels 78;	Gaps 10;
QY	1	TAGGATCCTTCAATGAAAATGTGTTATTTCTCTCATCACAG--ACAAAGGGCAACAGT	58	
DB	645	TAGGATCCTTCAATGAAAATGTGTTATTTCTCATCACAGGAAAAAAGGCAACAGT	704	
QY	59	TACAAAAACAATTTATGTTTCATTTGAGATTTAAGGAAGTTAAGGAAGAAAAAGATTA	118	
DB	705	TACACACAAAAATTTATGTTTCATTTGAGATTTAAGGAAGTTAAGGAAGAAAAAGATTA	764	
QY	119	AAAAATGTCCTATCTCTTTGTTT-----	143	
DB	765	AAAAATGTCCTATCTCTTTGTTATGCTCTTTTCGGAATTAATTAAGTTTAGTTGTA	824	
QY	144	-CTGTAATAATAATAAGAGACTTAAACCTTTTAATATAATAATGTAATTAGTTTT--	200	
DB	825	ATTGTAATAATAATAAGAGACTTAAACCTTTTAATAATAATAATTAATTAGTTTTTT	884	
QY	201	-----CTAGTCATGAGGACCACTCAGAGACAAGATTTCAGAGAAACAATTTTGT	250	
DB	885	TTAGTCACGAGTGTGATTCCTCACCCCTCAAAACAACATTTCCAGAGAAACAATTTGTT	944	
QY	251	AAACATCTTATTAGAAACTTTTAGTTTAAAGTCTTTGAAAGTTAGAAATAAACAAAAAAT-T	309	
DB	945	AAACATCTTATTAGAAACTTTTAGCAAGACTTTGAGTTAGAAATAAACAAAAAATTTAC	1004	
QY	310	ACACAGAGAAACACAATAAACCCACTACCGTCAGGTTATCATTAAGATGAATGTTTG	369	
DB	1005	ACACAGAGAAACACAATAAACCTACTATCGTCAGGTTATCATTAAGGATGAATGTTTG	1064	
QY	370	ATATCATTAATAATAACACACACAAAAATACATCTAATTAACATATATGTTATACAT	429	
DB	1065	ATACCATTAATAATAACACACAAAAAATACATGTAATTAACAAATACATGTTATAC--	1122	
QY	430	ATATTTTGTGAAAAACTTAGAGTTTTTCAAAACATTC-TAATACATGATTAGAGTTTATA	488	
DB	1123	ATATTTTGTGAAAAACTTAGAAGTTTTTCAAAACATTCCTTAATACATGATTAGAGTTTATA	1182	
QY	489	GAAATACAAATATTAAAAAATAATAATTTTAAAAAAAACATTCCTTAAGTCATTCAGATCCT	548	
DB	1183	GAAATACAAATATTAAAAAATAATACTTTTAAAAAAAACATCTTAAAGTC-CTCATATCCT	1241	
QY	549	CTCACACCTGTGTGATCATTTTAGTTCAT-GTATGTAGTACAATCAATCTGTAGTTCACACAG	607	
DB	1242	CTCACACCGGTGAAATCATTTTACTCTGTAGTATAGTACCGTGTCTAATAAGTTTCCACACAC	1301	
QY	608	AGTAAAAATAAATAGGATAAACTAGGGAATATATATAATATATACAATTAATAAAAAAG	667	
DB	1302	AGTAAAAAGAATAAAGATAAACTAGTGAATATA-----AAATTTTATAAAAAG	1349	
QY	668	GGAATATCAATTTAGAAATTTTGTATTTCCACATGACACACTCACCATCGAGCTGCCA	727	
DB	1350	GGAATATCAATTTAGAAATTTTGTATTTCCACATGACACACTCACCATGAGCTGCCA	1409	

KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.

OS Homo sapiens.

PN WO200277272-A2.

XX 03-OCT-2002.

PD 26-MAR-2002; 2002WO-EP003401.

PF 26-MAR-2001; 2001US-0278333P.

PR (EPIC-) EPIGENOMICS AG.

PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;

PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;

PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;

PI Schwobe I, Ziebarth H;

XX WPI; 2003-018942/01.

DR Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.

PS Claim 28; SEQ ID NO 339; 117pp; English.

XX The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ9861 to ABZ1118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used; for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferation disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients

XX SQ Sequence 3683 BP; 716 A; 0 C; 760 G; 2207 T; 0 U; 0 Other;

Query Match 7.6%; Score 87; DB 7; Length 3683;

Best Local Similarity 46.6%; Pred. No. 4.4e-06;

Matches 312; Conservative 0; Mismatches 355; Indels 2; Gaps 1;

QY 54 ACAGTTAAACAAACAAATTTATGTTTCATTTGAGATTAGGAGGTAGGAGAGAAAAAG 113

DB 3311 AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3252

QY 114 ATTAATAAAAAATGCTTATCTCTTTGTTTCTGTAATATATATAGAGACTTAAACTT 173

DB 3251 TAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3192

QY 174 TTAATATATAATTTGTAATTTAGTTTCTTCATGNGCACCACCTCAGAGCAAGATTTC 233

DB 3191 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3134

QY 234 AAAAAACAAATTTGTTTAAACATCTTATAGAACTTTTAGTCTTGTAGTTAGAA 293

DB 3133 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3074

QY 294 TTAACAAAAAATAATTAACACGAGAAACACAATAAACCCACTACCTCAGGTTATCAT 353
 DB 3073 TAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3014
 QY 354 AGGATGAATGTTTGTATATATCAATTAATAAACAACACAAAAATACATCTAATTATAAC 413
 DB 3013 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2954
 QY 414 AATATATGTTATACATATATTTTGTAAAAAATCTTAGAGTTTTCAAAAACATCTAATACA 473
 DB 2953 TAAACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2894
 QY 474 TGATTAGAGTTTATAGAAATACAAATATTTTAAAAAATAATATTTTAAAAAATAATTTCTAA 533
 DB 2893 AATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2834
 QY 534 AGTCATTAGATCTCTTCAACACCTGTGTGATCATTTTAGTGTATGTAGTACATCATTT 593
 DB 2833 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2774
 QY 594 GTAGTTCACACGAGTAAATAATAAATAGGATAAACTAGGGAATATATATAATATATACA 653
 DB 2773 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2714
 QY 654 ATTAAATAAAAAAGGGAAAAATCAATTTAGAAATTTTGAATTTCCCAATGACACAACTCAC 713
 DB 2713 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2654
 QY 714 CATGCACGC 722
 DB 2653 CTTTCCAC 2645

RESULT 9

ID ABK87142 standard; DNA; 4846 BP.

AC ABK87142;

DT 07-OCT-2002 (first entry)

DE Scarlet runner bean C541 genomic sequence.

XX Expression cassette; promoter activity; suspensor cell; plant embryo;
 KW modulation of gene transcription; Scarlet runner bean; C541; plant; gene;
 KW transgenic; ds.

OS Phaseolus coccineus.

PK Key Location/Qualifiers

FT CDS 3134..3552

FT /tag= a
 FT /product= "C541 protein"

PN WO200244333-A2.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-US044737.

XX 28-NOV-2000; 2000US-00724857.

XX 28-NOV-2000; 2000US-0253672P.

XX (REGC) UNIV CALIFORNIA.

XX (CERS-) CERES INC.

XX Weterings K, Apuya NR, Tatarinova T, Goldberg RB;

XX WPI; 2002-508506/54.

XX P-PSDB; RAU98691.

XX Expression cassette comprises promoters with basal promoter activity

PT operably linked to a heterologous polynucleotide, useful for expression
PT genes in suspensor cells in plants and/or basal region of plant embryo.
XX
PS Claim 25; Fig 4; 114pp; English.
XX
XX The present invention relates to expression cassettes comprising a
CC promoter sequence and a promoter polynucleotide with basal promoter
CC activity, where the promoter sequence is operably linked to a
CC heterologous polynucleotide, and when the expression cassette is inserted
CC into a plant, the heterologous polynucleotide is specifically expressed
CC in a suspensor cell and/or basal region of a plant embryo. The invention
CC also provides polynucleotide sequences encoding scarlet runner bean
CC (Phaseolus coccineus) G564 and C541 proteins for use in the expression
CC cassettes of the invention. The expression cassettes comprising promoters
CC and promoter control elements are useful for modulating transcription of
CC genes in a plant suspensor cell and/or basal region of a plant embryo.
CC The present sequence represents Scarlet runner bean C541 genomic sequence
XX
SQ Sequence 4846 BP; 1894 A; 777 C; 753 G; 1415 T; 0 U; 7 Other;
Query Match 7.4%; Score 85.2; DB 6; Length 4846;
Best Local Similarity 65.5%; Pred. No. 9.2e-06;
Matches 156; Conservative 0; Mismatches 78; Indels 4; Gaps 2;
QY 464 TTCTATACATGATTACAGTTTATAGAAATACAAATATTTAAATATATATTTAAAAA 523
DB 4086 TTATATTACAAATTTATGACTTATAGAAATACAAATATTTAAATATATA-AGGTTCAAAAC 4144
QY 524 AACATCTTAAGTCATTCAGATCTCTCACACCTGTGTATCATTTAGTCATGTATGTAG 583
DB 4145 TACATCTTAAGTCTTTTCAGACCCCTCTGACATGT---ATCATCTGCTGATATGTGA 4201
QY 584 TACAATCATTTAGTTTACACACGAGTAAATATAATAGGATATAAATAGGGAATATATAT 643
DB 4202 TACAGTCATCGCAGTTTACACAGATACACAGAAACCAAGGTAAGCTAATGAAAAAAT 4261
QY 644 AATATATACATTAATTAATAAAGGAAATCAATCAATAGATTTTGGATCCCAT 701
DB 4262 TCCATAACATATTTAATTCATGCAAAAGAACAGTCAGTAAAGTAATCATTTATAAAT 4319
RESULT 10
ABN83926
ID ABN83926 standard; DNA; 288 BP.
XX
AC ABN83926;
XX
DT 06-SEP-2002 (first entry)
XX
DE Arcelin-5 promoter sequence fragment.
XX
KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
XX WO200250295-A2.
XX
PD 27-JUN-2002.
XX
XX 17-DEC-2001; 2001WO-US047495.
XX
XX 18-DEC-2000; 2000US-0255879P.
XX
XX (RENE-) RENESSEN LLC.
XX
XX Wang Q, Dubois P, Liang J, Oulmassov T;
XX
XX WPI; 2002-508809/54.
XX
XX New transformed or transgenic soybeans plants or cells with an Arcelin-5
PT promoter, for use as an improved dietary source of protein for humans or
PT animals, or for producing soybeans with important agricultural or

PT nutritional properties.
XX
XX Example 1; Fig 1; 74pp; English.
XX
XX The invention relates to a transformed soybean plant cell and transgenic
CC soybean plant, both of which has a nucleic acid molecule comprising the
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
CC The transformed soybean plant cell and transgenic soybean plant are
CC useful as an improved source of dietary protein for humans and livestock.
CC These are also useful for producing soybean plants that exhibit important
CC agricultural, nutritional or pharmaceutical properties. The current
CC sequence represents an arcelin-5 promoter sequence fragment
XX
SQ Sequence 288 BP; 121 A; 26 C; 38 G; 103 T; 0 U; 0 Other;
Query Match 7.4%; Score 85; DB 6; Length 288;
Best Local Similarity 100.0%; Pred. No. 7.9e-06;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1037 TCGTTGTTTAAATTAATTTGAGTTTGAATTAATAATGAAAGAAAAAGTTGGAA 1096
DB 204 TCGTTGTTTAAATTAATTTGAGTTTGAATTAATAATGAAAGAAAAAGTTGGAA 263
QY 1097 AGATTTGCAATTTGTTTGTATATA 1121
DB 264 AGATTTGCAATTTGTTTGTATATA 288
RESULT 11
ABZ10100
ID ABZ10100 standard; DNA; 8056 BP.
XX
AC ABZ10100;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
XX Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
XX WO20027272-A2.
XX
PD 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-EP003401.
XX
XX 26-MAR-2001; 2001US-0278333P.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
XX Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
XX Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
XX Schwabe I, Ziebarth H;
XX
XX WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
XX Claim 28; SEQ ID NO 240; 117pp; English.
XX
XX The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG

CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 XX Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match 7.4%; Score 85; DB 7; Length 8056;
 Best Local Similarity 49.7%; Pred. No. 1.e-05;
 Matches 297; Conservative 0; Mismatches 295; Indels 5; Gaps 3;

QY 100 AAGGAGAGAAAAAGATTAABAAAATGCTTATCTCTTCTGTTCTGTAATAATAATA 159
 Db 1639 AATTAAAAAAATTTTAAATTAATACGATAAAATTTATTTTATTAATTAATAA 1698
 QY 160 AGAGACTTAAACTTTTATATA-ATAATGTGAATAGTGTTCCTAGTCATGAGCACC 218
 Db 1699 ATTAATTTAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1758
 QY 219 CAGAGACAGATTTCAAGAAACAATTTTGTAAACATCTTATAGAACTTTAGTTAA 278
 Db 1759 AAATTTTAAATTAATAAAATTTAAATTAATACGTTATTAATTTTAAATTAATA 1818
 QY 279 GTCTTGAAGTTAGAAATTAACACAAAAAAATTAACACAGAGAAACACATAACCCACTAC 338
 Db 1819 AATAAATAATTCGTTTAAATAAAAAAATAAAAAAATAAAAAAATAATTAATTAATAT 1878
 QY 339 CGTCAGGTTATCAAGAGATGAATGTTTGTGATATCATTAATAATAACACACAAAAAT 398
 Db 1879 TTTTATTTT- -TTATTTAATAAAATAAAAAATTTTATAAAAAAATAAAATTAATAT 1936
 QY 399 ACATCTAATTAACAAATATATGTTATACATATATTTTGTAAAAAATCTAGAGTTTCA 458
 Db 1937 AATATAATAAATAAATAATAATTTTAAATAAATAATAATAATAATAATAATAATA 1996
 QY 459 AAACATTTCTAATACATGATGAGTTTATGAAATACAAATATTTTAAAAAATAAATTT 518
 Db 1997 AATTTAATAAAATTTTAAATATATTTAAATAAATAATAATAATAATAATAATAATTT 2056
 QY 519 AAAAAACAATCT- -AAAGTCATTCAGATCCCTCACACCTGTGTGATCATTTAGTCATG 576
 Db 2057 TAAAAAATTTTATTAATAAATAATTTTAAATAATAATAATAATAATAATAATAATA 2116
 QY 577 TATGTAGTACATCATCTGTAGTTCAACAGAGTAAAAATAATAATAAGGATAAAGGAA 636
 Db 2117 TAAITTTAATAATTTTATTTTATTAATAAATAAATAAATAAATAAATAAATAATAAT 2176
 QY 637 TATATTAATAATATACAAATTAATAAATAAAGGAAAAATCAAAATTAAGATTTTGAAT 693
 Db 2177 TTTATTAATAAATTTTAAAAAATCGAAAAAATAAATAAATAAATAAATAAATAAATTT 2233

RESULT 12
 ABZ10246/C
 ID ABZ10246 standard; DNA; 8056 BP.
 XX
 AC ABZ10246;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.

XX Human; haematopoietic cell proliferation disorder; cytostatic;
 XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 XX cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200277272-A2.
 PD 03-OCT-2002.
 XX
 XX 26-MAR-2002; 2002WO-EP003401.
 PF
 XX 26-MAR-2001; 2001US-0278333P.
 PR
 XX (SPIG-) EPIGENOMICS AG.
 PA
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwabe I, Ziebarth E;
 XX WPI; 2003-018942/01.
 DR
 XX Detecting and differentiating between haematopoietic cell proliferative
 XX disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 PT
 XX Claim 28; SEQ ID NO 386; 117pp; English.
 PS
 XX The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match 7.4%; Score 84.8; DB 7; Length 8056;
 Best Local Similarity 48.7%; Pred. No. 1.e-05;
 Matches 290; Conservative 0; Mismatches 302; Indels 4; Gaps 2;

QY 71 TTTATGTTTTCATTTGAGATTAAGGAGGTAAGGAGAGAAAAAGATTAAAAAATGTCCT 130
 Db 2220 TTTTATTTTATTTTATTTTTCATTTTATTTTAAATTTTAAATAATTAATTTTATTTT 2161
 QY 131 TATCTCTTCTTCTGTAATAATAATAAGAGCTTAAACTTTTAAATAATAATTAATTA 190
 Db 2160 TTTTATTTTATTTTATTTTAAATAATAATAATAATAATAATAATAATAATAATAATA 2101
 QY 191 ATTAGGTTTCTTAGTCATGAGCACCCTCAGACAGAGATTTCAAGAAACAATTTGTT 250
 Db 2100 ATTATTTTATTAATAATTTTAAATAATTTTAAATAATAATTTTAAATAATAATAATTT 2041
 QY 251 AAACATCT-TAATAGAACTTTTAGTTAGCTTTAGCTTTAGAGTTAGAAATTAACAAAAAAT 309


```
Db 6675 TAAATAACACATTTCTCAAAAAAATAAATACAAATACCCACCAACATATAAAAAACACTCA 6616
Qy 425 TACATATATTTTGTAAAACTTAGAGTTTTCAAAAACATTTCTAAATACATGATGAGTT 484
Db 6615 ACATCACTTATTATCAAAAAAATAACAATCAAAACCCACAAATATAACCAATTTTACTCCT 6556
Qy 485 TATAGAAATACAATATTTTAAAAAATATAATTTTAAAAAAACATTTCTAAAAGTCATTGAGA 544
Db 6555 ACAAATAAACCATCATTAATAATAAAAAAATAAATAATTAATTAACATTAATATAATAAA 6496
Qy 545 TCCTCTCACACCTGTGTGATCATTTAGTCATGTAGTAGTACAAATCATTTGTAGTTGACAA 604
Db 6495 AAAAAACACTTTTACACTACCAATAAAAAATATAAACTAAATACCACTATAAAAAACAA 6436
Qy 605 CAGAGTAAATATAATAGGATAACTAGGGAATATATATATATATACAAATTAATAAAAA 664
Db 6435 CATAAAAACTAATAACAACCACTATAAACTAATACACCACTATAAAAAACATATAAAA 6376
Qy 665 AAGGGAATCAAAATTAGAATTT 687
Db 6375 AATTAAAAATAAACTACCAATTT 6353
```

Search completed: June 2, 2004, 02:24:19
Job time : 546 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 2, 2004, 02:26:24 ; Search time 4650 Seconds
(without alignments)
10700.603 Million cell updates/sec
Title: US-10-015-637-1
Perfect score: 1148
Sequence: 1 taggaccccttcataagaaaa.....acagagtgatggttaatgca 1148

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0
Searched: 3470272 seqs, 21671516995 residues
Word size : 0
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database :		GenEmbl:
1:	gb.ba.*	1: gb.ba.*
2:	gb.htg.*	2: gb.htg.*
3:	gb.in.*	3: gb.in.*
4:	gb.ov.*	4: gb.ov.*
5:	gb.pat.*	5: gb.pat.*
6:	gb.ph.*	6: gb.ph.*
7:	gb.pr.*	7: gb.pr.*
8:	gb.ro.*	8: gb.ro.*
9:	gb.sts.*	9: gb.sts.*
10:	gb.sy.*	10: gb.sy.*
11:	gb.un.*	11: gb.un.*
12:	gb.vi.*	12: gb.vi.*
13:	em.fun.*	13: em.fun.*
14:	em.in.*	14: em.in.*
15:	em.mu.*	15: em.mu.*
16:	em.ov.*	16: em.ov.*
17:	em.pat.*	17: em.pat.*
18:	em.ph.*	18: em.ph.*
19:	em.pl.*	19: em.pl.*
20:	em.ro.*	20: em.ro.*
21:	em.sts.*	21: em.sts.*
22:	em.un.*	22: em.un.*
23:	em.vi.*	23: em.vi.*
24:	em.htg_inv.*	24: em.htg_inv.*
25:	em.htg_other.*	25: em.htg_other.*
26:	em.htg_mus.*	26: em.htg_mus.*
27:	em.htg_pln.*	27: em.htg_pln.*
28:	em.htg_rod.*	28: em.htg_rod.*
29:	em.htg_man.*	29: em.htg_man.*
30:	em.htg_vrt.*	30: em.htg_vrt.*
31:	em_sy.*	31: em_sy.*
32:	em_htgo_hum.*	32: em_htgo_hum.*
33:	em_htgo_mus.*	33: em_htgo_mus.*
34:	em_htgo_other.*	34: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	1148	6	AX463281 Sequence
2	1148	100.0	1832	6	AX463294 Sequence
3	1122	97.7	1122	6	AX463282 Sequence
4	1046	91.1	1821	6	AX343309 Sequence
5	1046	91.1	3900	8	PVARC5X1
6	107	9.3	1866	6	AX463293 Sequence
7	107	9.3	1872	6	AX463292 Sequence
8	85	7.4	288	6	AX463291 Sequence
9	69	6.0	2288	8	AF193029 Phaseolus
10	65	5.7	1992	8	PVDFL3C2
11	49	4.3	1889	8	PVLECT
12	48	4.2	4564	8	PVARC1A
13	46	4.0	1768	8	PVDFLEC1
14	42	3.7	1852	8	AF255723
15	41	3.6	1211	6	E09281
16	41	3.6	1395	8	PVDFLEC1
17	41	3.6	1441	8	PVDFLEC1
18	27	2.4	68875	9	AC096359
19	27	2.4	126318	9	AC128689
20	27	2.4	153165	2	AC015539
21	27	2.4	171589	2	AC073567
22	27	2.4	175466	2	AC016419
23	27	2.4	204954	2	AC128028
24	27	2.4	276343	2	AC125570
25	26	2.3	316	6	AX463290
26	26	2.3	322	6	AX463289
27	26	2.3	1188	8	PVDFLEC2A
28	26	2.3	1661	8	PVDFLEC2
29	26	2.3	1784	8	PHYDLECB
30	24	2.1	32818	3	AC024825
31	24	2.1	148761	9	AL450352
32	24	2.1	150601	9	AC117426
33	24	2.1	160099	2	AC073207
34	24	2.1	174383	5	AL329216
35	24	2.1	184119	9	AC090348
36	24	2.1	185796	2	AP001398
37	24	2.1	187023	2	AC006882
38	24	2.1	193595	2	AC073681
39	24	2.1	234679	2	EX511263
40	24	2.1	251579	2	AC073677
41	23	2.0	13529	3	AY159038
42	23	2.0	40589	3	CEZK1025
43	23	2.0	40589	3	CEZK1025
44	23	2.0	46054	2	AC145619
45	23	2.0	66806	9	AP000022

ALIGNMENTS

RESULT 1	AX463281	AX463281	1148 bp	DNA	linear	PAT 15-JUL-2002
LOCUS	Sequence 1 from Patent WO0250295.					
DEFINITION	AX463281					
ACCESSION	AX463281.1	GI:21886232				
VERSION						
KEYWORDS	Phaseolus vulgaris					
SOURCE	Phaseolus vulgaris					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.					
REFERENCE	1					
AUTHORS	Culmassov, T., Wang, Q., Dubois, P. and Liang, J.					

```

TITLE Arcelin-5 promoter and uses thereof
JOURNAL Patent: WO 0250295-A 1 27-JUN-2002;
RENSSEN LLC (US)
FEATURES Location/Qualifiers
    source
        1..1148
            /organism="Phaseolus vulgaris"
            /mol_type="unassigned DNA"
            /db_xref="taxon:3885"
ORIGIN
    Query Match 100.0%; Score 1148; DB 6; Length 1148;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGATCCTTCATAGAAAATGTTATTTCTTCATCACCAGACAAAGGGGCAACAGTTA 60
DB 1 TAGGATCCTTCATAGAAAATGTTATTTCTTCATCACCAGACAAAGGGGCAACAGTTA 60
QY 61 ACAAAACAAATTTATGTTTCATTTGAGATTAAGGAGGTAAAGGAAAGAAAGATTAAA 120
DB 61 ACAAAACAAATTTATGTTTCATTTGAGATTAAGGAGGTAAAGGAAAGAAAGATTAAA 120
QY 121 AAAATGTCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 121 AAAATGTCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
QY 181 AATAATGTAATTAAGTTTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 240
DB 181 AATAATGTAATTAAGTTTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 240
QY 241 CAATTTGTTAAACATCTTATAGAACTTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTA 300
DB 241 CAATTTGTTAAACATCTTATAGAACTTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTA 300
QY 301 AAAAAAATTACACACGAGAAACACAAATAACCCACTACCGTCAGGTTATCATAGGATGA 360
DB 301 AAAAAAATTACACACGAGAAACACAAATAACCCACTACCGTCAGGTTATCATAGGATGA 360
QY 361 AATGTTTGTATATCATTAATATACACACACAAATAACCCACTACCGTCAGGTTATCAT 420
DB 361 AATGTTTGTATATCATTAATATACACACACAAATAACCCACTACCGTCAGGTTATCAT 420
QY 421 GTTATACATATATTTTGTAAAACTTTAGAGTTTTCCTCAAAACATTTCTAATACATGAT 480
DB 421 GTTATACATATATTTTGTAAAACTTTAGAGTTTTCCTCAAAACATTTCTAATACATG 480
QY 481 AGTTTATAGAAATACAAATATTTAAATAATTAATTTTAAATAATTAATTTTAAATAAT 540
DB 481 AGTTTATAGAAATACAAATATTTAAATAATTAATTTTAAATAATTAATTTTAAATAAT 540
QY 541 CAGATCCTCTCACACCTCTGTCATCTTATAGTCTAGTCTAGTCTAGTCTAGTCTAGTCT 600
DB 541 CAGATCCTCTCACACCTCTGTCATCTTATAGTCTAGTCTAGTCTAGTCTAGTCTAGTCT 600
QY 601 ACAACAGAGTAAATAAATTAAGGATTAACCTAGGGAATATATATATATATATATATAT 660
DB 601 ACAACAGAGTAAATAAATTAAGGATTAACCTAGGGAATATATATATATATATATATAT 660
QY 661 AAAAAAGGGAAATCAAAATAGAAATTTTTCATTTCCCAATGACACAACTCACCATGCAC 720
DB 661 AAAAAAGGGAAATCAAAATAGAAATTTTTCATTTCCCAATGACACAACTCACCATGCAC 720
QY 721 GCTGCCACCTCAGCTCCCTCTCTCCACATGTCATGTCATGTCATGTCATGTCATGTCAT 780
DB 721 GCTGCCACCTCAGCTCCCTCTCTCCACATGTCATGTCATGTCATGTCATGTCATGTCAT 780
QY 781 TTCACTATGACACAACTCGCCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 840
DB 781 TTCACTATGACACAACTCGCCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 840
QY 841 GACACCACTGGGATGATGCTGTCACCTCAGCTCCACCTCAGCTCCACCTCAGCTCCAC 900
DB 841 GACACCACTGGGATGATGCTGTCACCTCAGCTCCACCTCAGCTCCACCTCAGCTCCAC 900
QY 901 TGGCCATGCACACTGCCACCTCAGCACCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 960
DB 901 TGGCCATGCACACTGCCACCTCAGCACCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 960
QY 961 CTTCCTCTCCATTAATATCTATTTAAATTTAACTAAATTTAACTAAATTTAACTAAAT 1020
DB 961 CTTCCTCTCCATTAATATCTATTTAAATTTAACTAAATTTAACTAAATTTAACTAAAT 1020
QY 1021 CTTGATGTCATTTGCCATCTGTTTGTAAATTTGTTTAAATTTGTTTAAATTTGTTTAA 1080
DB 1021 CTTGATGTCATTTGCCATCTGTTTGTAAATTTGTTTAAATTTGTTTAAATTTGTTTAA 1080
QY 1081 AAAAAAAGTTGGAAGATTTTGCATTTGTTTGTATTAATATAGAGAGAGAGATGATGG 1140
DB 1081 AAAAAAAGTTGGAAGATTTTGCATTTGTTTGTATTAATATAGAGAGAGAGATGATGG 1140
QY 1141 TTAATGCA 1148
DB 1141 TTAATGCA 1148
RESULT 2
AX463294
LOCUS AX463294 1832 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 14 from Patent WO0250295.
ACCESSION AX463294
VERSION AX463294.1 GI:21886245
SOURCE Phaseolus vulgaris
ORGANISM Phaseolus vulgaris
REFERENCE 1. Culmasov, T., Wang, Q., Dubois, P. and Liang, J.
AUTHORS Arcelin-5 promoter and uses thereof
TITLE Patent: WO 0250295-A 14 27-JUN-2002;
JOURNAL RENESSEN LLC (US)
FEATURES Location/Qualifiers
    source
        1..1832
            /organism="Phaseolus vulgaris"
            /mol_type="unassigned DNA"
            /db_xref="taxon:3885"
ORIGIN
    Query Match 100.0%; Score 1148; DB 6; Length 1832;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 1148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGATCCTTCATAGAAAATGTTATTTCTTCATCACCAGACAAAGGGGCAACAGTTA 60
DB 659 TAGGATCCTTCATAGAAAATGTTATTTCTTCATCACCAGACAAAGGGGCAACAGTTA 718
QY 61 ACAAAACAAATTTATGTTTCATTTGAGATTAAGGAGGTAAAGGAAAGAAAGATTAAA 120
DB 719 ACAAAACAAATTTATGTTTCATTTGAGATTAAGGAGGTAAAGGAAAGAAAGATTAAA 778
QY 121 AAAATGTCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 779 AAAATGTCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 838
QY 181 AATAATGTAATTAAGTTTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 240
DB 839 AATAATGTAATTAAGTTTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 898
QY 241 CAATTTGTTAAACATCTTATAGAACTTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTA 300
DB 899 CAATTTGTTAAACATCTTATAGAACTTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTA 958
QY 301 AAAAAAATTACACACGAGAAACACAAATAACCCACTACCGTCAGGTTATCATAGGATGA 360
DB 301 AAAAAAATTACACACGAGAAACACAAATAACCCACTACCGTCAGGTTATCATAGGATGA 360
```

Db 959 AAAAAAATTACACACGAGAAACAAATAAACCCACTACCGTCAGGTTATCATBAAGATGA 1018
Qy 361 AATGTTTGGATATCATTAATAATAACACACACAAAAATACATCTAATATATAACATATAT 420
Db 1019 AATGTTTGGATATCATTAATAATAACACACACAAAAATACATCTAATATATAACATATAT 1078
Qy 421 GTTATACATATATTTTGTAAACCTTAGAGTTTTCACAAACATCTTAATACATGATTAG 480
Db 1079 GTTATACATATATTTTGTAAACCTTAGAGTTTTCACAAACATCTTAATACATGATTAG 1138
Qy 481 AGTTTATAGAAATACAAATATTTTAAATAATATATTTTAAATAAAACATCTTAAAGTCAT 540
Db 1139 AGTTTATAGAAATACAAATATTTTAAATAATATATTTTAAATAAAACATCTTAAAGTCAT 1198
Qy 541 CAGATCTCTCACACCTGTGTGATCATTTAGTCAATGATGATGATGATGATGATGATGATG 600
Db 1199 CAGATCTCTCACACCTGTGTGATCATTTAGTCAATGATGATGATGATGATGATGATGAT 1258
Qy 601 ACAACAGATGAAATAAATAAGGATAAATAGGATAAATAGGATAAATAAATAAATAAAT 660
Db 1259 ACAACAGATGAAATAAATAAGGATAAATAGGATAAATAAATAAATAAATAAAT 1318
Qy 661 AAAAAAGGAAATCAAAATTAAGATTTTGAATCCCAACATGACACAACTCACCATGAC 720
Db 1319 AAAAAAGGAAATCAAAATTAAGATTTTGAATCCCAACATGACACAACTCACCATGAC 1378
Qy 721 GCTGCCACCTCAGCTCCCTCTCTCCACACATGCTCATGCTCACTTTCGACTTTGGCTTT 780
Db 1379 GCTGCCACCTCAGCTCCCTCTCTCCACACATGCTCATGCTCACTTTCGACTTTGGCTTT 1438
Qy 781 TTCACTATGACACAACTCGCCATGATGTTGGCCACGAGCTCTCTCTCTCTCTCTCTCTCT 840
Db 1439 TTCACTATGACACAACTCGCCATGATGTTGGCCACGAGCTCTCTCTCTCTCTCTCTCTCT 1498
Qy 841 GACACCACTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 1499 GACACCACTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1558
Qy 901 TGCCATGACACACTGCCACCTCAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db 1559 TGCCATGACACACTGCCACCTCAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1618
Qy 961 CTCTCTCTCAATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
Db 1619 CTCTCTCTCAATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1678
Qy 1021 CGTCATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1679 CGTCATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1738
Qy 1081 AAAAAAAGTTGGAAGATTTGCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
Db 1739 AAAAAAAGTTGGAAGATTTGCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1798
Qy 1141 TTAATGCA 1148
Db 1799 TTAATGCA 1806

RESULT 3
AX463282
LOCUS AX463282 1122 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 2 from Patent WO0250295.
ACCESSION AX463282
VERSION AX463282.1 GI:2186233

KEYWORDS
SOURCE Phaseolus vulgaris
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1

AUTHORS Oulmasov, I., Wang, Q., Dubois, P. and Liang, J.

TITLE Arcelin-5 promoter and uses thereof

JOURNAL Patent: WO 0250295-A 2 27-JUN-2002;

RENESSEN LLC (US)

FEATURES Location/Qualifiers

source 1..1122

/organism="Phaseolus vulgaris"

/mol_type="unassigned DNA"

/db_xref="taxon:3885"

ORIGIN

Query Match 97.7%; Score 1122; DB 6; Length 1122;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 GGATCCTTCAATAGAAAATGCTTATTTCTCATCACCAGACAAAGGGGCAACAGTTAAC 62
Db 1 GGATCCTTCAATAGAAAATGCTTATTTCTCATCACCAGACAAAGGGGCAACAGTTAAC 60
Qy 63 AAAAAAATTTATTTGTTTCAATTTGAGATTAAAGGAAGTAAAGGAAGAAAAAGATTAAAAA 122
Db 61 AAAAAAATTTATTTGTTTCAATTTGAGATTAAAGGAAGTAAAGGAAGAAAAAGATTAAAAA 120
Qy 123 AATGCTCTTATCTCTTTGTTTCTGTAATAATAATATAGAGACTTAACTTTTAAATATA 182
Db 121 AATGCTCTTATCTCTTTGTTTCTGTAATAATAATAATAGAGACTTAACTTTTAAATATA 180
Qy 183 TAATTTGTAATTAAGTTTCTTAGTCATGAGCACCACCTCAGAGACAAAGATTTCAGAAAA 242
Db 181 TAATTTGTAATTAAGTTTCTTAGTCATGAGCACCACCTCAGAGACAAAGATTTCAGAAAA 240
Qy 243 ATTTGTTTAAACATCTTATTTAGAACTTTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAG 302
Db 241 ATTTGTTTAAACATCTTATTTAGAACTTTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAG 300
Qy 303 AAAAAATACACAGAGAAACACAAATAAACCCACTACCGTCAGGTTATCATAGAGTAA 362
Db 301 AAAAAATACACAGAGAAACACAAATAAACCCACTACCGTCAGGTTATCATAGAGTAA 360
Qy 363 TGTTTTGATATCATTAATAATAACACACACAAATAATACATCTAATATTAACATATATGT 422
Db 361 TGTTTTGATATCATTAATAATAACACACACAAATAATACATCTAATATTAACATATATGT 420
Qy 423 TATACATATATTTTCTGTAAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAG 482
Db 421 TATACATATATTTTCTGTAAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAG 480
Qy 483 TTTATAGAAATACAAATAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATA 542
Db 481 TTTATAGAAATACAAATAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATA 540
Qy 543 GATCCTCTCACACCTGTGTGATCATTTAGTCATGTTAGTTAGTTAGTTAGTTAGTTAGTT 602
Db 541 GATCCTCTCACACCTGTGTGATCATTTAGTCATGTTAGTTAGTTAGTTAGTTAGTTAGTT 600
Qy 603 AACAGAGTAAATAATAATAAGGATAAATAAGGATAAATAAGGATAAATAATAATAATAATA 662
Db 601 AACAGAGTAAATAATAATAAGGATAAATAAGGATAAATAATAATAATAATAATAATAATA 660
Qy 663 AAAAGGAAATCAAAATTAGAATTTTGAATCCCACTGACACAACTCAGCATGACCG 722
Db 661 AAAAGGAAATCAAAATTAGAATTTTGAATCCCACTGACACAACTCAGCATGACCG 720
Qy 723 TGCACCTCAGCTCCCTCTCTCCACACATGCTCTCATGTCACCTTTCGACTTTGGCTTTT 782
Db 721 TGCACCTCAGCTCCCTCTCTCTCCACACATGCTCTCATGTCACCTTTCGACTTTGGCTTTT 780
Qy 783 CACTATGACACAACTCGCCATGCAATGTTGCCAGTGAGCTCTCTCTCTCTCTCTCTCTCT 842
Db 781 CACTATGACACAACTCGCCATGCAATGTTGCCAGTGAGCTCTCTCTCTCTCTCTCTCTCT 840
Qy 843 CACCATGCGCATGCTGCTGCCACCTCAGTCCACCTCTCTCTCTCTCTCTCTCTCTCTCT 902

	/evidence=experimental	
3' UTR	2618..2754	
	/gene="arcs-I"	
	/citation=[2]	
	/citation=[1]	
polyA_signal	2704..2709	
	/gene="arcs-I"	
	/citation=[1]	
polyA_signal	2708..2713	
	/gene="arcs-I"	
polyA_signal	2722..2727	
	/gene="arcs-I"	
polyA_signal	2726..2731	
	/gene="arcs-I"	
ORIGIN		
Query Match	91.1%; Score 1046; DB 8; Length 3900;	
Best Local Similarity	99.8%; Pred. No. G;	
Matches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	1 TAGGATCCTTCAATAGAAATGTGTATTTCCTCATCCACAGACAAGGGGAACAGTTTA	60
DB		
QY	674 TAGGATCCTTCAATAGAAATGTGTATTTCCTCATCCACAGACAAGGGGGCAACAGTTTA	733
DB		
QY	61 ACATAACAAATTTATGTTTTCAITTCAGATTAAGGAAGTAAAGGAAGAAAAAGATTAAAA	120
DB		
QY	734 ACAAAACAAATTTATGTTTTCAITTCAGATTAAGGAAGTAAAGGAAGAAAAAGATTAAAA	793
DB		
QY	121 AAAATGTCCTTAATCTTTTGTTTCGTATAATATATAAGAGACTTAAACTTTTAAATAT	180
DB		
QY	794 AAAATGTCCTTAATCTTTTGTTTCGTATAATATATAAGAGACTTAAACTTTTAAATAT	853
DB		
QY	181 AATAATTTGTAATTAGGTTTTCTAGTCATGAGCACCACTCAGAGACAAGATTTCAAGAAA	240
DB		
QY	854 AATAATTTGTAATTAGGTTTTCTAGTCATGAGCACCACTCAGAGACAAGATTTCAAGAAA	913
DB		
QY	241 CAATTTTGTAAACATCTTTATTTAGAAAACTTTTAGTTTAAGTCTTGAAGTTAGAATTAAACA	300
DB		
QY	914 CAATTTTGTAAACATCTTTATTTAGAAAACTTTTAGTTTAAGTCTTGAAGTTAGAATTAAACA	973
DB		
QY	301 AAABAAAATTACACACAGAGAAACACAAATAAACCCACTACCCTCAGGTTATCATAAAGATGA	360
DB		
QY	974 AAAAAGAGTACACACAGAGAAACACAAATAAACCCACTACCCTCAGGTTATCATAAAGATGA	1033
DB		
QY	361 AATGTTTGTATATCAATTAATATAACACACACAAAAATACATCTAATTTATAACAATATAT	420
DB		
QY	1034 AATGTTTGTATATCAATTAATATAACACACACAAAAATACATCTAATTTATAACAATATAT	1093
DB		
QY	421 GTTATACATATATTTTGTAAAAACCTTAGAGTTTTTCAAAAACATTTCTAATATCATGATTAG	480
DB		
QY	1094 GTTATACATATATTTTGTAAAAACCTTAGAGTTTTTCAAACACATTTCTAATATCATGATTAG	1153
DB		
QY	481 AGTTTATAGAAATACAAATATTTTAAAAAATAATAATTTTAAAAAAAACATTTCTAAAGTCAT	540
DB		
QY	1154 AGTTTATAGAAATACAAATATTTTAAAAAATAATAATTTTAAAAAAAACATTTCTAAAGTCAT	1213
DB		
QY	541 CAGATCTCTCTCACACTGTGTGATCAATTTAGTCATGTATGTGTAGTACAATCATTTGTAGTTTC	600
DB		
QY	1214 CAGATCTCTCTCACACTGTGTGATCAATTTAGTCATGTATGTGTAGTACAATCATTTGTAGTTTC	1273
DB		
QY	601 ACACACAGAGTAAATAAATAAGATAAACAATAGGGAATATATATAATATATACAATTTAAAT	660
DB		
QY	1274 ACACACAGAGTAAATAAATAAGATAAACAATAGGGAATATATATAATATATACAATTTAAAT	1333
DB		
QY	661 AAAAAAGGGAATCAAAATTAGAAATTTTGTATCCCCCATGCACAACTCCACATGCCAC	720
DB		
QY	1334 AAAAAAGGGAATCAAAATTAGAAATTTTGTATCCCCCATGCACAACTCCACATGCCAC	1393
DB		
QY	721 GTTGGCACTCAGCTCCCTCTCTCCACACACATGTCTCATGTCACTTTCCGACTTTTGGCTTT	780
DB		
QY	1394 GTTGGCACTCAGCTCCCTCTCTCTCCACACATGTCTCATGTCACTTTCCGACTTTTGGCTTT	1453
DB		
QY	781 TTCACATATGACACAACTCGCCATGCATGTATGCCAGTGAGCTCTCTCTCTTCCCAATGAT	840

ACCESSION AF193029
 VERSION AF193029.1 GI:6684755
 KEYWORDS
 SOURCE Phaseolus vulgaris
 ORGANISM Phaseolus vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
 REFERENCE 1 (bases 1 to 2288)
 Gerhardt, I.R. and Grossi de Sa, M.F.
 TITLE Molecular characterization of a new arcelin-5 gene
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2288)
 Gerhardt, I.R.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1999) Embrapa/Cenargen, SAIN-Parque Rural-Final W5 Norte, Brasilia, DF 70770-900, Brazil
 FEATURES
 Location/Qualifiers
 1..2288
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /cultivar="G02771"
 /db_xref="taxon:3885"
 /clone="PIG53"
 /tissue_type="leaf"
 <603..>1385
 603..1385
 /note="arcs-III"
 /codon_start=1
 /product="arcelin 5c"
 /protein_id="AAP23725.1"
 /db_xref="GI:6684756"
 /translation="MASSNLLSLALFLVLLTHANSATETSFNFTSFHPDPMQLQSD ATISTKQLLTYSLESLGRALYSDFIQIKNNVASFDTKEFFIIRPETNGN SAYGLATLAVPGSKPKGKPYLGIFNDITPEPDARTAVVFNLRIDIDVNAIKP YANKECPHKYNGKTDVQIYDSSKNDLRVLFHTVSQVKCSVSATVQLEKEWENV SVGSATSGLTZENTETHDVLNWSFSKFRNKLNNLL"
 mRNA
 CDS
 Query Match 6.0%; Score 69; DB 8; Length 2288;
 Best Local Similarity 100.0%; Pred. No. 7.7e-24;
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 965 TCTCCATAAATATCTATTAAATTAACCTAACTAATTTATATATATCTTTTGTGACGCTG 1024
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 399 TCTCCATAAATATCTATTAAATTAACCTAACTAATTTATATATATCTTTTGTGACGCTG 458
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1025 GATGCATTTG 1033
 Db ||||||||
 Qy 459 GATGCATTTG 467
 RESULT 10
 PVPDLEC2
 LOCUS PVPDLEC2 1992 bp DNA linear PLN 21-MAR-1995
 DEFINITION P. vulgaris cv. Pinto Pdlec2 gene for phytohemagglutinin (PHA).
 ACCESSION X04659
 VERSION X04659.1 GI:21022
 KEYWORDS glycoprotein; Pdlec2 gene; phytohemagglutinin; phytohemagglutinin-L; signal peptide.
 SOURCE Phaseolus vulgaris
 ORGANISM Phaseolus vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
 REFERENCE 1 (bases 1 to 1992)
 Voelker, T.A., Staswick, P. and Chrispeels, M.J.
 TITLE Molecular analysis of two phytohemagglutinin genes and their expression in Phaseolus vulgaris cv. Pinto, a lectin-deficient cultivar of the bean

JOURNAL EMO J. 5, 3075-3082 (1986)
 COMMENT See also <X04660> for dlec1 gene (pseudogene). Several inverted repeats are described in the 5' upstream region of the PHA coding region.
 2 pot. glycosylation sites are found at the AA residues 33 and 92. Pdlec2 mRNA is reduced 20-fold (with respect to mRNA levels in normal cultivars); a 100bp deletion between pos. 501 and 502 (compared to dlec1, dlec2 and Pdlec1) containing a large tandem repeat may be responsible for the low level of expression of Pdlec2.
 FEATURES
 Location/Qualifiers
 1..1992
 /organism="Phaseolus vulgaris"
 /mol_type="genomic DNA"
 /strain="Pinto U111"
 /db_xref="taxon:3885"
 /tissue_type="young leaves (greenhouse)."
 742..747
 /note="put. TATA-box"
 785..1606
 /note="unnamed protein product; precursor polypeptide (AA -21 to 252)."
 /codon_start=1
 /protein_id="CAA28362.1"
 /db_xref="GI:21023"
 /db_xref="GOA:P15231"
 /db_xref="SWISS-PROT:P15231"
 /translations="MASSNLLSLALFLVLLTHANSASQTFFSDFRNETNLLIQGDAS VSSGQLRTNVNSNGEPTVSGLRAPFIQIMDTTIGNVASFDNTFFNLLVPPNN AGPADGLAFALVPGSQPKGKGLFGDGSNPHFVAVEFDLKNKMDPRRHIG IDVNSIKSIKTTTFDFVNGENAEVHITYSTKLLVASLYPSLKTSTFTVSDIVDLKS VLPWVSUGVGSATTTGTTKGNVETNLSWFSKLSGDTTSEGLNLANLVLNQIL"
 785..847
 /note="put. signal peptide (AA -21 to -1)"
 848..1603
 /product="mature PHA (AA 1-252)"
 1691..1699
 /note="put. polyA signal"
 ORIGIN
 Query Match 5.7%; Score 65; DB 8; Length 1992;
 Best Local Similarity 100.0%; Pred. No. 8.9e-22;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1084 AAAAAAGTTGGAAGATTTTGCATTTTGTGTTTGTATAAATAGAGAAGAGTGTATGTTA 1143
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 709 AAAAAAGTTGGAAGATTTTGCATTTTGTGTTTGTATAAATAGAGAAGAGTGTATGTTA 768
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1144 ATGCA 1148
 Db |||||
 Qy 769 ATGCA 773
 RESULT 11
 PHVLECT
 LOCUS PHVLECT 1689 bp DNA linear PLN 27-APR-1993
 DEFINITION P. vulgaris lectin gene, complete cds.
 ACCESSION J01261
 VERSION J01261.1 GI:169354
 KEYWORDS lectin.
 SOURCE Phaseolus vulgaris
 ORGANISM Phaseolus vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
 REFERENCE 1 (bases 257 to 1115)
 Hoffman, I.M., Ma, Y. and Barker, R.F.
 TITLE Molecular cloning of Phaseolus vulgaris lectin mRNA and use of cDNA as a probe to estimate lectin transcript levels in various tissues
 JOURNAL Nucleic Acids Res. 10 (23), 7819-7828 (1982)
 MEDLINE 83116994
 PUBMED 6897567

```

REFERENCE
AUTHORS      Hoffman,L.M.
TITLE        Structure of a chromosomal Phaseolus vulgaris lectin gene and its
              transcript
JOURNAL      J. Mol. Appl. Genet. 2 (5), 447-453 (1984)
MEDLINE      85008540
PUBMED       6090563
COMMENT      Original source text: P.vulgaris cv. Tendergreen DNA, clone
              p-lambda-lec5.7 [2]; and cotyledon, cDNA to mRNA, clone pPVL134
              [1].
CAT and TATA boxes are located at positions 165-169 and 226-231
              respectively. Two tandem polyadenylation signals are present at
              position 1084-1093.
              A magnetic tape of this sequence and a draft entry were kindly
              provided by L.M.Hoffman (23-OCT-1985).
FEATURES
source       Location/Qualifiers
              1..1689
               /organism="Phaseolus vulgaris"
               /mol_type="genomic DNA"
               /db_xref="taxon:3885"
mRNA         255..1115
               /product="lectin mRNA"
CDS          265..1005
               /note="lectin prepeptide"
               /codon_start=1
               /protein_id="AAA33769.1"
               /db_xref="GI:169355"
               /translation="MIMASSKILSLALPLALLSHANSATSFIIIDAFNKNLILQGD
               ATVSNGLQISYNSYDMSRAFYSAPIQIRDSITGNVASPTDFTNIRTHRCANSA
               VGLDFVLPVOPESKGDVTVEPTFLSRISIDNNNDIKSPVQVHDYDQNAEVEI
               TYNSKTRFVSLSVLSNPSTGKSNVSTTVELEKEVDVWVGVSATSGAYQWSYTHDV
               LSWFSFKFINKLQOKSERSNIVLNKIL"
sig_peptide 265..333
               /note="lectin signal peptide"
mat_peptide 334..1002
               /product="lectin mature peptide"
ORIGIN       71 bp upstream AhalII site.
              Query Match      4.3%; Score 49; DB 8; Length 1689;
              Best Local Similarity 100.0%; Pred. No. 1.5e-13;
              Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1100 TTTTCATTGTGTGTATTAATAGAGAGAGAGCTGTTAATGCA 1148
      |||||||||||||||||||||||||||||||||||||||||||
DB 209 TTTTCATTGTGTGTATTAATAGAGAGAGAGCTGTTAATGCA 257

RESULT 12
PHVARC1A    PHVARC1A      4564 bp DNA linear PLN 27-APR-1993
LOCUS       Phaseolus vulgaris arcelin (arc) gene, complete cds.
DEFINITION  M68913
VERSION     M68913.1 GI:169313
KEYWORDS   arcelin; lectin-like seed protein.
SOURCE     Phaseolus vulgaris
ORGANISM   Phaseolus vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1 (bases 1 to 4564)
AUTHORS    Anthony,J.L., Vonder Haar,R.A. and Hall,T.C.
TITLE      Nucleotide sequence of a genomic clone encoding arcelin, a
            lectin-like seed protein from Phaseolus
JOURNAL    Plant Physiol. 97, 839-840 (1991)
COMMENT    Original source text: Phaseolus vulgaris (library: Lambda zap II)
            young plant leaf DNA.
FEATURES
source       Location/Qualifiers
              1..4564
               /organism="Phaseolus vulgaris"
               /mol_type="genomic DNA"
               /db_xref="taxon:3885"
               /tissue_type="leaf"

/dev_stage="young plant"
/tissue_lib="Lambda Zap II"
3332..4172
/genes="arc"
3332..3338
/genes="arc"
3361..3364
/genes="arc"
/notes="cap region"
3375..4172
/genes="arc"
/codon_start=1
/product="arcelin"
/protein_id="AAA33753.1"
/db_xref="GI:169314"
/translation="MASSNLLTALPLVLLTHANSNDASFVETFNKTNLILQGDAT
VSSEGHLLNVKNGNEEDSGRAFYSAPIQIRDSITGNVASPTDFTNIRTHRCANSA
AYGLAFALVPVOPESKGLKRYLGFNTINRDRAHTVAVFDVTSNRIEDVNSIRPI
ATGSCNFGNNGEKAERVITYDSPKNDLRVSLSPSECKCHVSATVPLEKEVEDWVS
VGFSATSGSKKETTETHNVLSWFSFSSNFINKGKKSERSNIVLNKIL"
3375..3437
/genes="arc"
3438..4169
/genes="arc"
/product="arcelin"

ORIGIN
Query Match      4.2%; Score 48; DB 8; Length 4564;
Best Local Similarity 100.0%; Pred. No. 4.6e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1068 AATATAAATGAAGAAAAAGTTGGAAGATTTGCATTGTGTT 1115
      |||||||||||||||||||||||||||||||||||||||||||
DB 3283 AATATAAATGAAGAAAAAGTTGGAAGATTTGCATTGTGTT 3330

RESULT 13
PVPDLEC1    PVPDLEC1      1768 bp DNA linear PLN 12-SEP-1993
LOCUS       P. vulgaris cv.Pinto pseudogene Pdlec1 for phytohemagglutinin
DEFINITION  (PHA).
ACCESSION   X04660
VERSION     X04660.1 GI:21020
KEYWORDS   Pdlec1 gene; phytohemagglutinin; pseudogene.
SOURCE     Phaseolus vulgaris
ORGANISM   Phaseolus vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1 (bases 1 to 1768)
AUTHORS    Voelker,T.A., Staswick,P. and Chrispeels,M.J.
TITLE      Molecular analysis of two phytohemagglutinin genes and their
            expression in Phaseolus vulgaris cv. Pinto, a lectin-deficient
            cultivar of the bean
JOURNAL    EMBO J. 5, 3075-3082 (1986)
COMMENT    See also <X04659> for Pdlec2 gene.
            Several inverted repeats are described in the 5' upstream region of
            the PHA gene.
            None of the 3 possible reading frames allow the translation of its
            message into a complete PHA polypeptide. The PHA translation
            start codon (pos. 708-710) starts a RF which codes for a truncated
            50AA polypeptide with a N-terminal sequence completely identical to
            the dlec1 gene product, but is out of register after 10 codons due
            to a single bp deletion (pos. 739/740).
FEATURES
source       Location/Qualifiers
              1..1768
               /organism="Phaseolus vulgaris"
               /mol_type="genomic DNA"
               /strain="Pinto U111"
               /db_xref="taxon:3885"
               /tissue_type="young leaves (greenhouse)."
              665..670
               promoter

```

CDS /note="put. TATA-box"
708. .866
/note="unnamed protein product; pseudogene region"
/codon_start=1
/protein_id="CAA28363.1"
/db_xref="GI:21021"
/db_xref="REMBEM:CAA28363"
/translation="MASSNLSLALSLCFSPQTQPAKPFSSASIGSTKPTLSSKAMPP
SHPPASVD"
739. .740
variation /note="tt is tct in pDlec2 (c deletion, causing a frame
shift mutation)"
misc_feature 1620. .1635
/note="put. polyA signal"
ORIGIN
Query Match 4.0%; Score 46; DB 8; Length 1768;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1103 TGCATTGTTGTTGTTATAATAGAGAGAGAGTGATGGTTAATGCA 1148
DB 651 TGCATTGTTGTTGTTATAATAGAGAGAGAGTGATGGTTAATGCA 696
RESULT 14
AF255723 1852 bp DNA linear PLN 24-MAY-2000
LOCUS Phaseolus vulgaris ARL5-IV pseudogene, complete sequence.
DEFINITION AF255723
ACCESSION AF255723.1 GI:8050712
VERSION
KEYWORDS
SOURCE Phaseolus vulgaris
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1 (bases 1 to 1852)
AUTHORS Gerhardt, I.R., Pappas, G., Chrispeels, M.J. and Grossi de Sa, M.F.
TITLE Direct Submission
JOURNAL Submitted (13-APR-2000) Biotechnology, Embrapa-Cenargen, SAIN
Parque Rural- Final W5 Norte, CP 02372, Brasilia, DF 70 770-900,
Brazil
FEATURES
source Location/Qualifiers
1. .1852
/organism="Phaseolus vulgaris"
/mol_type="genomic DNA"
/db_xref="taxon:3885"
146. .814
/gene="ARL5-IV"
/pseudo
gene
ORIGIN
Query Match 3.7%; Score 42; DB 8; Length 1852;
Best Local Similarity 100.0%; Pred. No. 6.1e-10;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1098 GATTTGCATTTCTGTTGTTATAATAGAGAGAGTGATG 1139
DB 107 GATTTGCATTTCTGTTGTTATAATAGAGAGAGTGATG 148
RESULT 15
E09281 1211 bp DNA linear PAT 29-SEP-1997
LOCUS
DEFINITION DNA encoding lectin-like protein of kidney bean.
ACCESSION E09281
VERSION E09281.1 GI:22025907
KEYWORDS JP 1995132092-A/1.
SOURCE unidentified
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 1211)
AUTHORS Hagiwara, K.
TITLE NEW KIDNEY BEAN GENE
JOURNAL Patent: JP 1995132092-A 1 23-MAY-1995;
NORIN SUIJANSIO NOGIO SEIBUTSU SHIGEN KENKYUSHO
COMMENT Phaseolus sp. (kidney bean)
PN JP 1995132092-A/1
PD 23-MAY-1995
PI HAGIWARA KIYOSHI
PC C12N15/09, C12N1/21, C07K14/42, (C12N1/21, C12R1:19); CC
strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH 1. .1211
FT source /organism="Phaseolus sp."
FT /strain="Kentucky wonder".
FEATURES
source Location/Qualifiers
1. .1211
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 3.6%; Score 41; DB 6; Length 1211;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1108 TTGTTGTTGTTATAATAGAGAGAGTGATGGTTAATGCA 1148
DB 263 TTGTTGTTGTTATAATAGAGAGAGTGATGGTTAATGCA 303
Search completed: June 2, 2004, 06:17:34
Job time : 4660 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 03:34:13 ; Search time 108 Seconds
(without alignments)
5898.922 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148

Sequence: 1 taggacccctcaatagaaaa.....agagagtgtgtaataatgca 1148

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.8	7430	4	US-08-976-259-64 Sequence 64, Appl
2	20	1.7	802	4	US-09-807-258-19 Sequence 19, Appl
3	20	1.7	1597	4	US-09-071-035-87 Sequence 87, Appl
4	20	1.7	1677	4	US-09-134-000C-2844 Sequence 2844, Ap
5	20	1.7	1680	4	US-09-071-035-85 Sequence 85, Appl
6	20	1.7	3627	1	US-08-104-072B-6 Sequence 6, Appl
7	20	1.7	3627	1	US-08-351-413-7 Sequence 7, Appl
8	20	1.7	3627	2	US-09-025-583-7 Sequence 7, Appl
9	20	1.7	392000	4	US-10-027-983-11 Sequence 11, Appl
10	20	1.7	640681	4	US-09-790-988-1 Sequence 1, Appl
11	20	1.7	1230025	4	US-09-198-452A-1 Sequence 1, Appl
12	19	1.7	181	4	US-09-669-751-213 Sequence 213, App
13	19	1.7	1212	4	US-09-489-039A-2871 Sequence 2871, Ap
14	19	1.7	1776	4	US-09-134-001C-391 Sequence 391, App
15	19	1.7	2590	4	US-09-023-655-301 Sequence 301, App
16	19	1.7	4105	4	US-09-634-238-182 Sequence 182, App
17	19	1.7	4784	4	US-09-634-238-24 Sequence 24, Appl
18	19	1.7	30246	4	US-08-956-171E-56 Sequence 56, Appl
19	19	1.7	38653	4	US-09-922-445-1 Sequence 1, Appl
20	19	1.7	319608	4	US-09-539-333D-1 Sequence 1, Appl
21	19	1.7	319608	4	US-09-679-409-1 Sequence 1, Appl
22	19	1.7	640681	4	US-09-790-988-1 Sequence 1, Appl
23	18	1.6	109	4	US-09-539-601-45 Sequence 45, Appl
24	18	1.6	115	4	US-09-539-601-50 Sequence 50, Appl
25	18	1.6	115	4	US-09-539-601-51 Sequence 51, Appl
26	18	1.6	300	4	US-08-956-171E-3111 Sequence 3111, Ap
27	18	1.6	311	4	US-09-539-401C-59 Sequence 59, Appl

28 18 1.6 424 4 US-09-643-597-157 Sequence 157, App
29 18 1.6 424 4 US-09-480-884A-157 Sequence 157, App
30 18 1.6 424 4 US-09-542-615A-157 Sequence 157, App
31 18 1.6 424 4 US-09-606-421B-157 Sequence 157, App
32 18 1.6 424 4 US-09-221-107-157 Sequence 157, App
33 18 1.6 424 4 US-09-640-173-157 Sequence 176, App
34 18 1.6 487 4 US-09-713-550-176 Sequence 176, App
35 18 1.6 487 4 US-09-556-002-15 Sequence 15, Appl
36 18 1.6 615 4 US-08-956-171B-937 Sequence 937, App
37 18 1.6 816 4 US-09-134-001C-1926 Sequence 1926, Ap
38 18 1.6 916 3 US-09-239-909-3 Sequence 3, Appl
39 18 1.6 927 4 US-09-598-401C-29 Sequence 29, Appl
40 18 1.6 1001 4 US-09-641-638-160 Sequence 160, App
41 18 1.6 1001 4 US-09-641-638-161 Sequence 161, App
42 18 1.6 1072 1 US-07-971-096-1 Sequence 1, Appl
43 18 1.6 1072 1 US-08-175-096-1 Sequence 1, Appl
44 18 1.6 1172 4 US-09-598-401C-90 Sequence 90, Appl
45 18 1.6 1200 3 US-09-327-681-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-976-259-64
; Sequence 64, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7430 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-976-259-64

Query Match 1.8%; Score 21; DB 4; Length 7430;
Best Local Similarity 100.0%; Pred.No. 2.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 500 ATTAAAAATATAATTAA 520
|||||

Db 7289 ATTAAATAATATAATTTAA 7309

RESULT 2

US-09-807-258-19

Sequence 19, Application US/09807258

Patent No. 6670166

GENERAL INFORMATION:

APPLICANT: E. I. du Pont de Nemours and Company

TITLE OF INVENTION: Arthropod Protein Disulfide Isomerases

FILE REFERENCE: BB-1253 PCT

CURRENT APPLICATION NUMBER: US/09/807,258

CURRENT FILING DATE: 2001-06-11

PRIOR APPLICATION NUMBER: 60/104,376

PRIOR FILING DATE: 1998-10-15

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Microsoft Office 97

SEQ ID NO 19

LENGTH: 802

TYPE: DNA

ORGANISM: Hottentotta judiaca

US-09-807-258-19

Query Match 1.7%; Score 20; DB 4; Length 802;

Best Local Similarity 100.0%; Pred. No. 8.3; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1070 TAATAAAATCAAGAAAGAAA 1089

Db 771 TAATAAAATCAAGAAAGAAA 790

RESULT 3

US-09-071-035-87

Sequence 87, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 1597 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-87

Query Match 1.7%; Score 20; DB 4; Length 1597;

Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 AAAATCAAATTAGAAATTTT 689

Db 1112 AAAATCAAATTAGAAATTTT 1131

RESULT 4

US-09-134-000C-2844

Sequence 2844, Application US/09134000C

Patent No. 6617156

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-032

CURRENT APPLICATION NUMBER: US/09/134,000C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/055,778

PRIOR FILING DATE: 1997-08-15

NUMBER OF SEQ ID NOS: 6812

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2844

LENGTH: 1677

TYPE: DNA

ORGANISM: Enterococcus faecalis

US-09-134-000C-2844

Query Match 1.7%; Score 20; DB 4; Length 1677;

Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 AAAATCAAATTAGAAATTTT 689

Db 1189 AAAATCAAATTAGAAATTTT 1208

RESULT 5

US-09-071-035-85

Sequence 85, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 1597 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-87

```
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-071-035-85

Query Match          1.7%; Score 20; DB 4; Length 1680;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 670 AAAATCAATAGATTATT 639
Db 1192 AAAATCAATAGATTATT 1211

RESULT 6
US-08-104-072B-6/c
; Sequence 6, Application US/08104072B
; Patent No. 5639948
; GENERAL INFORMATION:
; APPLICANT: Michiels, Frank
; APPLICANT: Moriooka, Sinji
; APPLICANT: Scheirlinck, Trees
; APPLICANT: Komari, Toshiko
; TITLE OF INVENTION: Stamen-specific Promoters from Rice
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5639948 West Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/104,072B
; FILING DATE: 05-AUG-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 9200272
; FILING DATE: 06-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91403352.7
; FILING DATE: 10-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91402590.3
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91400318.1
; FILING DATE: 08-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.93USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORGANISM: Oryza sativa

; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..2845
; OTHER INFORMATION: /function= "sequence comprising
; OTHER INFORMATION: anther-specific PT72 promoter".
; FEATURE:
; NAME/KEY: TATA signal
; LOCATION: 2733..2739
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2765
; OTHER INFORMATION: /product= "transcription
; OTHER INFORMATION: initiation"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2846
; OTHER INFORMATION: /product= "ATG start translation of
; OTHER INFORMATION: T72 gene"
US-08-104-072B-6

Query Match          1.7%; Score 20; DB 1; Length 3627;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1071 AATAAATGAAGAAAAAG 1090
Db 991 AATAAATGAAGAAAAAG 972

RESULT 7
US-08-351-413-7/c
; Sequence 7, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSES: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2045
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Oryza sativa
STRAIN: Akihikari
FEATURE:
NAME/KEY: -
LOCATION: 1..2845
OTHER INFORMATION: /label= PT72
OTHER INFORMATION: /note= "sequence comprising anther specific
OTHER INFORMATION: promoter PT72"
FEATURE:
NAME/KEY: -
LOCATION: 2733..2739
OTHER INFORMATION: /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: -
LOCATION: 2765
OTHER INFORMATION: /note= "transcription initiation
OTHER INFORMATION: determined by primer extension"
FEATURE:
NAME/KEY: -
LOCATION: 2846..2848
OTHER INFORMATION: /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"
US-08-351-413-7

Query Match 1.7%; Score 20; DB 1; Length 3627;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1071 AATAAATGAAGAAAAAG 1090
|||||
Db 991 AATAAATGAAGAAAAAG 972

RESULT 8
US-09-025-583-7/c
Sequence 7, Application US/09025583
Patent No. 5977433
GENERAL INFORMATION:
APPLICANT: Williams, Mark
APPLICANT: Leemans, Jan
TITLE OF INVENTION: Maintenance of male-sterile plants
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 2046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/025,583
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/351,413
FILING DATE:
APPLICATION NUMBER: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:

NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3627 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Oryza sativa
STRAIN: Akihikari
FEATURE:
NAME/KEY: -
LOCATION: 1..2845
OTHER INFORMATION: /label= PT72
OTHER INFORMATION: /note= "sequence comprising anther specific
OTHER INFORMATION: promoter PT72"
FEATURE:
NAME/KEY: -
LOCATION: 2733..2739
OTHER INFORMATION: /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: -
LOCATION: 2765
OTHER INFORMATION: /note= "transcription initiation
OTHER INFORMATION: determined by primer extension"
FEATURE:
NAME/KEY: -
LOCATION: 2846..2848
OTHER INFORMATION: /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"
US-09-025-583-7
Query Match 1.7%; Score 20; DB 2; Length 3627;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1071 AATAAATGAAGAAAAAG 1090
|||||
Db 991 AATAAATGAAGAAAAAG 972
RESULT 9
US-10-027-983-11/c
Sequence 11, Application US/10027983
Patent No. 8617162
GENERAL INFORMATION:
APPLICANT: Kenneth W. Dobie
APPLICANT: Mark P. Roach
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
FILE REFERENCE: RTS-0340
CURRENT APPLICATION NUMBER: US/10/027,983
CURRENT FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 11
LENGTH: 392000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 137740
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 137742


```

; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (138122)...(138221)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 145507
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 151967
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (151967)...(1542066)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 154217
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (164037)...(164136)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (174657)...(174756)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (186224)...(186323)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (195242)...(195341)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 202703
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (202771)...(202870)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (206246)...(215602)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (218126)...(218225)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (220360)...(220459)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (222717)...(222816)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (223581)...(224080)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (227487)...(227586)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (230157)...(230256)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (232299)...(232398)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (236552)...(236651)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (238789)...(248788)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: exon
; LOCATION: (118288)...(119101)
; OTHER INFORMATION: exon 1C
; NAME/KEY: exon:intron junction
; LOCATION: (151129)...(151130)
; OTHER INFORMATION: exon 5:intron 5
; NAME/KEY: exon:intron junction
; LOCATION: (299248)...(299249)
; OTHER INFORMATION: exon 9:intron 9

```

```

; NAME/KEY: exon:intron junction
; LOCATION: (348578)...(348579)
; OTHER INFORMATION: exon 10:intron 10
; NAME/KEY: intron
; LOCATION: (348579)...(381838)
; OTHER INFORMATION: intron 10
; NAME/KEY: intron:exon junction
; LOCATION: (386185)...(386186)
; OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

Query Match 1.7%; Score 20; DB 4; Length 392000;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1071 AATAAAATGAAAGAAAAAG 1090
DB 276309 AATAAAATGAAAGAAAAAG 276290

RESULT 10
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 1.7%; Score 20; DB 4; Length 640681;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 TATTTAAAAATATATTTT 518
DB 186362 TATTTAAAAATATATTTT 186381

RESULT 11
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(15000)
; OTHER INFORMATION: n=a or c or g or t

```

[illegible]

OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature

Query Match 1.7%; Score 20; DB 4; Length 1230025;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 GAAGAAAAAGATTAAAAA 122
DB 1175674 GAAGAAAAAGATTAAAAA 1175655

RESULT 12
US-09-669-751-213
Sequence 213, Application US/09569751
Patent No. 6551575
GENERAL INFORMATION:
APPLICANT: Greenspan, Ralph J.
TITLE OF INVENTION: Methods for Identifying Compounds for Motion Sickness, Vertigo and Other Disorders Related to Balance and the Perception of Gravity
FILE REFERENCE: P-NI 3864
CURRENT APPLICATION NUMBER: US/09/669,751
CURRENT FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US 60/168,579
PRIOR FILING DATE: 1999-12-02
NUMBER OF SEQ ID NOS: 261
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 213
LENGTH: 181
TYPE: DNA
ORGANISM: Drosophila
US-09-669-751-213

Query Match 1.7%; Score 19; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 AAATATAACACACAAAA 396
|||||

DB 107 AAATATAACACACAAAA 125
RESULT 13
US-09-489-039A-2871/c
Sequence 2871, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 2871
LENGTH: 1212
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2871

Query Match 1.7%; Score 19; DB 4; Length 1212;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 ACAATTAAATAAAAAAGGG 669
|||||
DB 47 ACAATTAAATAAAAAAGGG 29

RESULT 14
US-09-134-001C-391/c
Sequence 391, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 391
LENGTH: 1776
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-391

Query Match 1.7%; Score 19; DB 4; Length 1776;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 TTAATATATATATTTGTAAT 192
|||||
DB 547 TTAATATATATATTTGTAAT 529

RESULT 15
US-09-023-655-301
Sequence 301, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE EXPRESSION

NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/05/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 301:
SEQUENCE CHARACTERISTICS:
LENGTH: 2590 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LATRUT02
CLONE: 1353065
US-09-023-655-301

Query Match 1.7%; Score 19; DB 4; Length 2590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1068 AATAATAAATGAAGAAA 1086
|||||
DB 2265 AATAATAAATGAAGAAA 2283

Search completed: June 2, 2004, 07:16:20
Job time : 117 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:36:30 ; Search time 3393 Seconds
(without alignments)
10133.545 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148
Sequence: 1 taggtcttcctcaagaaaa.....agagagtatggttaatgca 1148

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513299 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	25	2.2	846	28	BH570165 BOGPE54TR
c 2	24	2.1	194	29	CE710533 tigr-gss-
c 3	24	2.1	491	29	BX122318 Danio rer
c 4	24	2.1	536	13	BX734218 BX734218

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	24	2.1	696	28	BZ882476
6	24	2.1	733	13	BX693278
7	24	2.1	820	29	CC474046
8	24	2.1	1509	29	CG755596
9	23	2.0	259	14	R95629
10	23	2.0	307	9	AA380112
11	23	2.0	449	14	CB378526
12	23	2.0	543	29	TA191802P
13	23	2.0	604	29	CE607569
14	23	2.0	611	29	CE223942
15	23	2.0	654	14	CB935043
16	23	2.0	747	29	CC586297
17	23	2.0	786	28	BZ163794
18	22	1.9	53	9	A1561282
19	22	1.9	162	14	TL14440
20	22	1.9	168	9	AA528923
21	22	1.9	171	9	AA525589
22	22	1.9	199	14	CD072757
23	22	1.9	201	9	AW017393
24	22	1.9	210	9	AA801591
25	22	1.9	210	9	A1977469
26	22	1.9	218	9	AA508323
27	22	1.9	221	9	AA801590
28	22	1.9	225	28	BH439146
29	22	1.9	240	14	R95407
30	22	1.9	245	9	AA528868
31	22	1.9	246	9	A1977349
32	22	1.9	250	14	R95403
33	22	1.9	300	14	R95405
34	22	1.9	306	9	A1976001
35	22	1.9	308	9	A1977237
36	22	1.9	310	14	R95404
37	22	1.9	326	14	CD171263
38	22	1.9	345	14	R95607
39	22	1.9	350	9	A1976293
40	22	1.9	359	14	CA905893
41	22	1.9	360	14	CD190460
42	22	1.9	368	9	AW017025
43	22	1.9	370	9	A1067007
44	22	1.9	372	9	A1976918
45	22	1.9	372	9	AW017201

ALIGNMENTS

RESULT 1
BH570165/c
LOCUS
DEFINITION
BOGPE54TR BOGP Brassica oleracea genomic clone BOGPE54, genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TIGR
JOURNAL
COMMENT

BH570165 846 bp DNA linear GSS 14-DEC-2001
BOGPE54TR BOGP Brassica oleracea genomic clone BOGPE54, genomic survey sequence.
BH570165 1 GI:17822019
GSS.
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 846)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOGPE54TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

```

FEATURES
  source
    Location/Qualifiers
      1..846
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="T0100DH3"
        /db_xref="taxon:3712"
        /clone="BOGPE54"
        /clone_lib="BOGP"
      /note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

ORIGIN
  Query Match      2.2%; Score 25; DB 28; Length 846;
  Best Local Similarity 100.0%; Pred. No. 3.4;
  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 ATTAAAAAATATAATTTAAAAAA 524
    |||||
Db 832 ATTAAAAAATATAATTTAAAAAA 808

RESULT 2
CE710533/c
LOCUS
DEFINITION
  tigr-gss-dog-I7000369461433 Dog Library Canis familiaris genomic,
  genomic survey sequence.
ACCESSION
  CE710533
VERSION
  CE710533.1 GI:37029944
KEYWORDS
  GSS:
SOURCE
  Canis familiaris (dog)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
  1 (bases 1 to 194)
  Kirknes,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
  Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
  Venter,J.C.
  The dog genome: survey sequencing and comparative analysis
  Science 301 (5641), 1898-1903 (2003)
  MEDLINE
  22875432
  PUBMED
  14512627
  COMMENT
    Contact: Kirknes EF
    The Institute for Genomic Research
    Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
    Rockville, MD 20850, USA
    Tel: 301-838-0200
    Fax: 301-838-0208
    Email: ekirknes@tigr.org
    Class: shotgun.
FEATURES
  source
    Location/Qualifiers
      1..194
        /organism="Canis familiaris"
        /mol_type="genomic DNA"
        /strain="Standard Poodle"
        /db_xref="taxon:9615"
        /clone_lib="Dog Library"
      /note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN
  Query Match      2.1%; Score 24; DB 29; Length 194;
  Best Local Similarity 100.0%; Pred. No. 13;
  Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 TATAACAATATATGTTATACATAT 431
    |||||
Db 155 TATAACAATATATGTTATACATAT 132

RESULT 3
BX122318
LOCUS
DEFINITION
  Danio rerio genomic clone DKEY-72P11, genomic survey sequence.
FEATURES
  source
    Location/Qualifiers
      1..491
        /organism="Danio rerio"
        /mol_type="genomic DNA"
        /db_xref="taxon:7955"
        /clone="DKEY-72P11"
        /tissue_type="testis"
        /note="vector pindigoBAC-536"

ORIGIN
  Query Match      2.1%; Score 24; DB 29; Length 491;
  Best Local Similarity 100.0%; Pred. No. 11;
  Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 CAATATTTAAAAAATAATTTT 518
    |||||
Db 89 CAATATTTAAAAAATAATTTT 112

RESULT 4
BX734218
LOCUS
DEFINITION
  BX734218 XGC-tadpole Silurana tropicalis cDNA clone TTPA076k17 5',
  mRNA sequence.
ACCESSION
  BX734218
VERSION
  BX734218.1 GI:38406959
KEYWORDS
  EST.
SOURCE
  Silurana tropicalis (western clawed frog)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
    Xenopodinae; Silurana.
    1 (bases 1 to 536)
REFERENCE
  Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
  Sanger Xenopus tropicalis EST project 2001 (11_2003)
  Unpublished (2003)
  TITLE
    JOURNAL
  COMMENT
    Contact: Croning MDR
    Sanger Institute
    Hinxton, Cambridgeshire, CB10 1SA, UK
    Email: trop@sanger.ac.uk
    Sanger Xenopus tropicalis EST project 2001
    TROPICALIS_SEQUENCE_ID: TTPA076k17.plkSP6
    Sequencing primer: SP6
    This sequence is from a Xenopus Gene Collection (XGC) library
    constructed by Nigel Garrett.
    cDNA was oligo dt primed from 5ug of poly A+ RNA from tadpole
    embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
    EcoRI at the 5' end and NotI at the 3' end.
    Vector: pCS107; Site1: EcoRI; Site2: NotI
    Host: Escherichia coli DH10B.
FEATURES
  source
    Location/Qualifiers
      1..536
        /organism="Silurana tropicalis"
        /mol_type="mRNA"
        /db_xref="taxon:8364"

```



```

/clone="TtpA076k17"
/dev stage="tadpole (stage 35-40)"
/lab_host="E. coli DH10B"
/clone_lib="XGC-tadpole"
/notes="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN
Query Match      2.1%; Score 24; DB 13; Length 536;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 630 TAGGGAATATATATATATATACA 653
Db 160 TAGGGAATATATATATATATACA 183

RESULT 5
BZ882476
LOCUS
DEFINITION
CH240_254K15_TV CHORI-240 Bos taurus genomic clone CH240_254K15,
genomic survey sequence.
ACCESSION
BZ882476
VERSION
BZ882476.1 GI:29109878
KEYWORDS
SOURCE
Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 696)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Chen,D., Riggs,P., de Jong,P.,
Crawford,A.M. and McGowan,J.C.
Bovine BAC End Sequences from Library CHORI-240
Unpublished (2003)
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.choi.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by AgResearch Ltd., New Zealand and The
Institute of Genomic Research (TIGR), USA.
Plate: 254 row: K column: 15
Seg primer: T7
Class: BAC ends.
Location/Qualifiers
1..696
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_254K15"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/notes="vector: pIARBAC1.3; Site_1: MboI; Site_2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match      2.1%; Score 24; DB 28; Length 696;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 ATAACACACACAAAATACATCTA 405
Db 351 ATAACACACACAAAATACATCTA 374

RESULT 6
BX693278/c
LOCUS
DEFINITION
BX693278 XGC-neurula Silurana tropicalis cDNA clone TNeu03f10 3',
mRNA sequence.
ACCESSION
BX693278
VERSION
BX693278.1 GI:38342398
KEYWORDS
SOURCE
EST.
ORGANISM
Silurana tropicalis (western clawed frog)
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
REFERENCE
1 (bases 1 to 733)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
JOURNAL
COMMENT
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu03f10.q1kT7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli DH10B.
Location/Qualifiers
1..733
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu03f10"
/dev stage="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match      2.1%; Score 24; DB 13; Length 733;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 630 TAGGGAATATATATATATACA 653
Db 448 TAGGGAATATATATATATACA 425

RESULT 7
CC474046
LOCUS
DEFINITION
CH240_298D10_TARBAC13P2 CHORI-240 Bos taurus genomic clone
CH240_298D10, genomic survey sequence.
ACCESSION
CC474046
VERSION
CC474046.1 GI:31751163
KEYWORDS
SOURCE
GSS.
Bos taurus (cow)
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


genes). The size range of inserts for the library was 0.5Kb - 2.5 Kb, the cDNA being directionally cloned between the EcoRI and the XhoI sites of the Bluescript phagemid. Bluescript phagemid was excised by mass in vivo excision of the library and inserted into E.coli XL1 Blue. The plasmid was purified using the Qiagen Qiaquick 8 system and sequenced with Taq dye-terminator chemistry (SK,T3 and T7 primers) for ABI 373A. The library was prepared by Hanan Abdel Hamed, Aïm Shams Univ., Cairo and provided by Mohamed Saber, Theodore Bilharz Research Institute, Cairo."

ORIGIN

Query Match 2.0%; Score 23; DB 14; Length 259;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AATATATATATATATATACAAATTA 657

DB 36 AATATATATATATATATACAAATTA 58

RESULT 10
 AA380112/c 307 bp mRNA linear EST 21-APR-1997
 LOCUS
 DEFINITION EST93059 Skin tumor I Homo sapiens cDNA 5' end, mRNA sequence.

AA380112

VERSION AA380112.1 GI:2032451

EST.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 307)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,I.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodsk,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Lib,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,B.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-P., Ferrie,A.A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wilg,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

7566098

Other ESTs: THC110532

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018693056

Fax: 3018693423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse

Location/Qualifiers

1. 307

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="ATCC (inhost):184517"

FEATURES

source

/db_xref="taxon:9606"

/dev_stages="adult"

/clone_lib="Skin tumor I"

/note="Organ: skin; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI"

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 307;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1085 AAAAAGTTGGAAAGATTTCAT 1107

DB 268 AAAAAGTTGGAAAGATTTCAT 246

RESULT 11

CB378526

LOCUS

DEFINITION

CB378526

ACCESSION

CB378526

VERSION

CB378526.1

KEYWORDS

CB378526

SOURCE

CB378526

ORGANISM

DEFINITION

CB378526

REFERENCE

AUTHORS

McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,

Wyllie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,

Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,

Tsagarisvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,

Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,

Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,

McCann,R., Waterston,R. and Wilson,R.

The Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This library was generated by cloning cDNAs directionally into

Uni-ZAP(Stratagene) (T3 primer/EcoRI are at the 5'-end and T7/XhoI

are at the 3'-end). The library was excised [now in pBluescript

SK(+)] and normalized (Bonaldo et al 1996 Genome Research

6:791-806). Library constructed by Thomas Baum (tbaum@iastate.edu),

Iowa State University, Plant Pathology Department and Jeff

McDermott (jmcderm@iastate.edu).

Putative full length read

The vector to vector length is 479

Seq primer: T3 from Gibco

High quality sequence stop: 449.

Location/Qualifiers

1. 449

/organism="Heterodera glycines"

/mol_type="mRNA"

/db_xref="taxon:51029"

/sex="mixed"

/tissue_type="whole organism"

/dev_stages="3rd stage juvenile"

/lab_host="DH10B"

/clone_lib="Heterodera glycines J3"

/note="Vector: pBluescript SK+ (Stratagene); Site 1: XhoI;

Site 2: EcoRI; This library was generated by cloning cDNAs

directionally into Uni-ZAP(Stratagene) (T3 primer/EcoRI

are at the 5'-end and T7/XhoI are at the 3'-end). The

library was excised [now in pBluescript SK(+)] and

normalized (Bonaldo et al 1996 Genome Research 6:791-806).

Library constructed by Thomas Baum (tbaum@iastate.edu),


```
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature

Query Match 1.7%; Score 20; DB 4; Length 1230025;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 GAAGAAAAAGATTAATAAAA 122
Db 1175674 GAAGAAAAAGATTAATAAAA 1175655

RESULT 12
US-09-669-751-213
; Sequence 213, Application US/09669751
; Patent No. 6551575
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Methods for Identifying Compounds for
; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
; TITLE OF INVENTION: Balance and the Perception of Gravity
; FILE REFERENCES: P-NI 3864
; CURRENT APPLICATION NUMBER: US/09/669,751
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 60/168,579
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 261
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 181
; TYPE: DNA
; ORGANISM: Drosophila
US-09-669-751-213

Query Match 1.7%; Score 19; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 AAATATAACACACACAAAA 396
|||||

OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature

Query Match 1.7%; Score 19; DB 4; Length 1212;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 ACAATTAAATAATAAAAGGG 669
|||||
Db 47 ACAATTAAATAATAAAAGGG 29

RESULT 14
US-09-134-001C-391/c
; Sequence 391, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCES: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 391
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-391

Query Match 1.7%; Score 19; DB 4; Length 1776;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 TTAATATAATAATTGTAAT 192
|||||
Db 547 TTAATATAATAATTGTAAT 529

RESULT 15
US-09-023-655-301
; Sequence 301, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
```

```

NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 301:
SEQUENCE CHARACTERISTICS:
LENGTH: 2590 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LAINUT02
CLONE: 1353065
US-09-023-655-301

```

```

Query Match      1.7%; Score 19; DB 4; Length 2590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1068 AATAATAAATGAAGAAA 1086
    |||||
Db 2265 AATAATAAATGAAGAAA 2283

```

```

Search completed: June 2, 2004, 07:16:20
Job time : 117 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:24:29 ; Search time 537 Seconds
(without alignments)
9081.809 Million cell updates/sec

Title: US-10-015-637-1

Perfect score: 1148

Sequence: 1 taggattccttaataagaaaa.....agagagtgtggttaatgca 1148

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_28Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1148	100.0	1148	ABN83916	ABN83916 Truncated
2	1148	100.0	1832	ABN83929	ABN83929 Arcelin-5
3	1122	97.7	1122	ABN83917	ABN83917 Truncated
4	1046	91.1	1821	AA229066	AA229066 Phaseolus
5	107	9.3	1866	ABN83928	ABN83928 Arcelin-4
6	107	9.3	1872	ABN83927	ABN83927 Arcelin-3
7	85	7.4	288	ABN83926	ABN83926 Arcelin-5
8	41	2.6	1211	AA294051	AA294051 Kidney be
9	26	2.3	316	ABN83925	ABN83925 Arcelin-4
10	26	2.3	322	ABN83924	ABN83924 Arcelin-3
11	25	2.2	1542	AA290025	AA290025 DNA encod
12	24	2.1	24	AB280122	AB280122 Arcelin 5
13	24	2.1	1224	AA262538	AA262538 cDNA sequ
14	23	2.0	400	AB46475	AB46475 Human bre
15	23	2.0	400	AAK30701	AAK30701 Human bon
16	23	2.0	400	AAK05134	AAK05134 Human bra
17	23	2.0	5067	AD48876	AD48876 Human gen
18	22	1.9	469	AA243419	AA243419 Human G-P
19	22	1.9	2453	AB21668	AB21668 Drosophi
20	22	1.9	7374	ABK40036	ABK40036 Human che
21	22	1.9	7374	AB170602	AB170602 Chemical
22	22	1.9	10080	AA295933	AA295933 Human KIK
23	22	1.9	14312	AAK87515	AAK87515 Human imm

24	22	1.9	14313	4	AAK87517	Human imm
25	22	1.9	14321	4	AAK87516	Human imm
26	21	1.8	2890	7	AB235868	Human sec
27	21	1.8	3161	7	ACC46534	Human dit
28	21	1.8	5237	6	ABL34617	Human met
29	21	1.8	5982	6	ABL32180	Human imm
30	21	1.8	7430	2	AAV31250	E. coli J
31	21	1.8	8662	6	ABL34636	Human met
32	21	1.8	9504	6	ABK28408	DNA trans
33	21	1.8	9760	6	ABK31242	Signal tr
34	21	1.8	9760	6	ABL70197	Chemical
35	21	1.8	9760	6	AA611155	Human gen
36	21	1.8	12409	6	AA63312	Chemical
37	21	1.8	16633	6	ABN79984	Human Che
38	20	1.7	210	4	AAI25078	Probe #15
39	20	1.7	210	4	ABA70751	Human fce
40	20	1.7	210	4	AAI50924	Probe #19
41	20	1.7	210	4	ABA37251	Probe #15
42	20	1.7	210	4	AAK44954	Human bon
43	20	1.7	210	4	AAK19004	Human bra
44	20	1.7	210	4	ABS44619	Human liv
45	20	1.7	210	6	ABS19198	Human gen

ALIGNMENTS

RESULT 1

ABN83916
ID ABN83916 standard; DNA; 1148 BP.
XX
AC ABN83916;
XX
DT 06-SEP-2002 (first entry)
DE Truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
XX
KW Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
PN WO200250295-A2.
XX
PD 27-JUN-2002.
XX
PF 17-DEC-2001; 2001WO-US047495.
XX
PR 18-DEC-2000; 2000US-0255879P.
XX (RENE-) RENESSEN LLC.

Wang Q, Dubois P, Liang J, Oulmasov T;

WPI; 2002-508809/54.

New transformed or transgenic soybeans plants or cells with an Arcelin-5 promoter, for use as an improved dietary source of protein for humans or animals, or for producing soybeans with important agricultural or nutritional properties.

Claim 4; Page 65-66; 74pp; English.

The invention relates to a transformed soybean plant cell and transgenic soybean plant, both of which has a nucleic acid molecule comprising the Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence. The transformed soybean plant cell and transgenic soybean plant are useful as an improved source of dietary protein for humans and livestock. These are also useful for producing soybean plants that exhibit important agricultural, nutritional or pharmaceutical properties. The current sequence represents the truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter sequence. This sequence is a deletion mutant of the Arcelin-5 promoter, created by the removal of approximately 600 base

719	DB	ACAAAA	CAAAATTTATGTTCTTCATTTTGAGATTAAAGGAGGTTAAGGAGAAAAAGAGTTAAAA	778
121	QY	AAAATGCTCTTATCTCTTTTGTTTCTGTAATAATAATAAGAGACTTAAAC	TTTAAATAT	180
779	DB	AAAATGCTCTTATCTCTTTTGTTTCTGTAATAATAATAAGAGACTTAAAC	TTTAAATAT	838
181	QY	AATAATTTGAATTTAGGTTTCTAGTCATGAGCAACCACTCAGAGACAAGATTTCAAGAAA	240	
839	DB	AATAATTTGAATTTAGGTTTCTAGTCATGAGCAACCACTCAGAGACAAGATTTCAAGAAA	898	
241	QY	CAATTTTGTTTAAACATCTTTATTAGAAAATCTTTTAGTTAAAGTCTTGAAGTTAGAACTAAACA	300	
899	DB	CAATTTTGTTTAAACATCTTTATTAGAAAATCTTTTAGTTAAAGTCTTGAAGTTAGAACTAAACA	958	
301	QY	AAAAAAATTTACACAGAGAAAACAATAAACCACTACCGTCAGGTTATCATTAAGATGA	360	
959	DB	AAAAAAATTTACACAGAGAAAACAATAAACCACTACCGTCAGGTTATCATTAAGATGA	1018	
361	QY	AATGTTTTCATATCATTTAAATATAACACACAAAAATACATCTAATATTATAACAATATAT	420	
1019	DB	AATGTTTTCATATCATTTAAATATAACACACAAAAATACATCTAATATTATAACAATATAT	1078	
421	QY	GTTATACATATATTTTGTAAAAAATCTTAGAGTTTTCAAAAATCTCTAATATACATGAATTAG	480	
1079	DB	GTTATACATATATTTTGTAAAAAATCTTAGAGTTTTCAAAAATCTCTAATATACATGAATTAG	1138	
481	QY	AGTTTATAGAAATACAAAATATTTTAAAAATATATATTTTAAAAAAAACATTTCTAAAGTCATT	540	
1139	DB	AGTTTATAGAAATACAAAATATTTTAAAAATATATATTTTAAAAAAAACATTTCTAAAGTCATT	1198	
541	QY	CAGATCCTCTCACACCTGTGTGATCATTTTAGTCATGTATAGTACAAATCAATTTGTAGTTTC	600	
1199	DB	CAGATCCTCTCACACCTGTGTGATCATTTTAGTCATGTATAGTACAAATCAATTTGTAGTTTC	1258	
601	QY	ACACAGAGTTAAATATAATAAGAGNTAAATCTTAGGGAATATATAATAATATACAAATTAAT	660	
1259	DB	ACACAGAGTTAAATATAATAAGAGNTAAATCTTAGGGAATATATAATAATATACAAATTAAT	1318	
661	QY	AAAAAAGGAAAAATCAAAATPAGAAATTTTTCATTTCCCAATGACACAACTCACCATGCGAC	720	
1319	DB	AAAAAAGGAAAAATCAAAATPAGAAATTTTTCATTTCCCAATGACACAACTCACCATGCGAC	1378	
721	QY	GCTGCCACCTCAGCTCCCTCTCTCCACATGTCTCATGTGTCACTTTTCGACTTTGGCTTT	780	
1379	DB	GCTGCCACCTCAGCTCCCTCTCTCCACATGTCTCATGTGTCACTTTTCGACTTTGGCTTT	1438	
781	QY	TTCACTATGACACAACTCGCCATGTTGCCACGAGAGCTCTCTCTCTCTCCCATGAT	840	
1439	DB	TTCACTATGACACAACTCGCCATGTTGCCACGAGAGCTCTCTCTCTCTCCCATGAT	1498	
841	QY	GACACCACTGGGCGATGCTGCTGCCACCTCAGCTCCCACTCTCTCTCTCAATATGAGCGCTAC	900	
1499	DB	GACACCACTGGGCGATGCTGCTGCCACCTCAGCTCCCACTCTCTCTCTCAATATGAGCGCTAC	1558	
901	QY	TGCGCCATGACACCTGCGCCACCTCAGCTCCCTCTCTCTCTCCCATGCTGCTGCCAAACCG	960	
1559	DB	TGCGCCATGACACCTGCGCCACCTCAGCTCCCTCTCTCTCTCCCATGCTGCTGCCAAACCG	1618	
961	QY	CTTCTCTCCATAAATATCTATTTAAATTTAAACCTAATATTATTCATATCTTTTTCATGA	1020	
1619	DB	CTTCTCTCCATAAATATCTATTTAAATTTAAACCTAATATTATTCATATCTTTTTCATGA	1678	
1021	QY	CGTGGATGCAATGCCATCGTGTGTTTAATAATTTGTAATTTGGAGTTGCAATAATAAATGA	1080	
1679	DB	CGTGGATGCAATGCCATCGTGTGTTTAATAATTTGTAATTTGGAGTTGCAATAATAAATGA	1738	
1081	QY	AAGAAAAAAGCTTCGAAAGATTTTTCATTTCTGTTGTGTATATAATAGAGAGAGAGTGTATGG	1140	
1739	DB	AAGAAAAAAGCTTCGAAAGATTTTTCATTTCTGTTGTGTATATAATAGAGAGAGAGTGTATGG	1798	
1141	QY	TTAATGCA	1148	
1799	DB	TTAATGCA	1806	

Query Match	Best Local Similarity	Score	DB	Length
ABN83917 standard; DNA; 1122 BP.	97.7%	1122	DB 6;	Length 1122;
ABN83917;	100.0%;	Pred. NO. 0;	Mismatches 0;	Indels 0; Gaps 0;
06-SEP-2002 (first entry)				
Truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter #2.				
Arcelin-5; promoter; plant; transgenic; soybean; agriculture; nutrition;				
pharmaceutical; ds.				
Phaseolus vulgaris.				
WO200250295-A2.				
27-JUN-2002.				
17-DEC-2001; 2001WO-US047495.				
18-DEC-2000; 2000US-0255879P.				
(RENE-) RENESSEN LLC.				
Wang Q, Dubois P, Liang J, Oulmassov T;				
WPI; 2002-508809/54.				
New transformed or transgenic soybeans plants or cells with an Arcelin-5 promoter, for use as an improved dietary source of protein for humans or animals, or for producing soybeans with important agricultural or nutritional properties.				
Example 5; Page 66; 74pp; English.				
The invention relates to a transformed soybean plant cell and transgenic soybean plant, both of which has a nucleic acid molecule comprising the Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence. The transformed soybean plant cell and transgenic soybean plant are useful as an improved source of dietary protein for humans and livestock. These are also useful for producing soybean plants that exhibit important agricultural, nutritional or pharmaceutical properties. The current sequence represents a truncated P. vulgaris exotic genotype G02771 Arcelin-5 promoter sequence. This sequence was used in the expression of a GUS reporter gene in a soybean cotyledon				
Sequence 1122 BP; 410 A; 210 C; 146 G; 356 T; 0 U; 0 Other;				
Query Match	97.7%	Score 1122	DB 6;	Length 1122;
Best Local Similarity	100.0%;	Pred. NO. 0;	Mismatches 0;	Indels 0; Gaps 0;
Matches 1122; Conservative				
GGATCTCTCAATAGAAAATGTGTATTTCCTCATCAGACAAAAGGGGCAACAGTTAAC 62				
1 GGATCTCTCAATAGAAAATGTGTATTTCCTCATCAGACAAAAGGGGCAACAGTTAAC 60				
63 AAAACAAATTTATGTTTCATTTGGATTTAGGAGTTAAGGAGAAAAGATTAAAAAA 122				
61 AAAACAAATTTATGTTTCATTTGGATTTAGGAGTTAAGGAGAAAAGATTAAAAAA 120				
123 AATGTCCTTATCTCTGTTCTGTAATAAATATATATAGAGACTTAAACCTTTTATATAA 182				
121 AATGTCCTTATCTCTGTTCTGTAATAAATATATATAGAGACTTAAACCTTTTATATAA 180				
183 TAATTTGTAATTAGGTTTCTAGTCAATGAGCACCCTCAGAGACAGAGATTTTCAGAGAAAACA 242				
181 TAATTTGTAATTAGGTTTCTAGTCAATGAGCACCCTCAGAGACAGAGATTTTCAGAGAAAACA 240				
243 ATTTTGGTTAAACAATCTTATTAGAACTTTTAGTTTAACTTTTGAAGTTAGATTATTAACAAA 302				

241	Db	ATTTTGTAAACATCTTATTAGAAACCTTTAGTTAGCTTGTGAAGTTAGTAATTAACAAA	300
303	Qy	AAAAATTACACACGAGGAAACACAAATAAACCCACTACCGTCAGGTTATCATPAAGATGAAA	362
301	Db	AAAAATTACACACGAGGAAACACAAATAAACCCACTACCGTCAGGTTATCATPAAGATGAAA	360
363	Qy	TGTTTTGATATCAATTAATAATAACACACACAAAAAATACATCTAAATTATTAACAATATATGT	422
361	Db	TGTTTTGATATCAATTAATAATAACACACACAAAAAATACATCTAAATTATTAACAATATATGT	420
423	Qy	TATACATATATTTTGTAAAAACCTTAGAGTTTTTCAAAAACATTCCTAATCATCATTAGAG	482
421	Db	TATACATATATTTTGTAAAAACCTTAGAGTTTTTCAAAAACATTCCTAATCATCATTAGAG	480
483	Qy	TTTATAGAAATACAAATATTTAAAAAATAATAATTTTAAAAAAAACATTCCTAAAGCTATTCA	542
481	Db	TTTATAGAAATACAAATATTTAAAAAATAATAATTTTAAAAAAAACATTCCTAAAGCTATTCA	540
543	Qy	GATCCTCTCACACTGTGTGATCACTTTAGTCATGTATGTAGTACATCATTTGTAGTTCAC	602
541	Db	GATCCTCTCACACTGTGTGATCACTTTAGTCATGTATGTAGTACATCATTTGTAGTTCAC	600
603	Qy	AACAGAGTAAAAATAATPAGGATAAACTAGGCAATATATATATATATACAAATTAATAAA	662
601	Db	AACAGAGTAAAAATAATPAGGATAAACTAGGCAATATATATATATATACAAATTAATAAA	660
663	Qy	AAAAAGGAAAACTAAATTAGAAATTTTGGATTTCCCAACATGACACAACTGACCATGCACGC	722
661	Db	AAAAAGGAAAACTAAATTAGAAATTTTGGATTTCCCAACATGACACAACTGACCATGCACGC	720
723	Qy	TGCAACCTCAGCTCCCTCCCTCTCCACACATGTCTCATGTCACTTTGCACTTTGGCTTTTT	782
721	Db	TGCAACCTCAGCTCCCTCCCTCTCCACACATGTCTCATGTCACTTTGCACTTTGGCTTTTT	780
783	Qy	CACATGACACAACTGCCATGCAATTTGGATTTCCCAACATGACACAACTGACCATGCATGA	842
781	Db	CACATGACACAACTGCCATGCAATTTGGATTTCCCAACATGACACAACTGACCATGCATGA	840
843	Qy	CACCACTGGGCACTGCATGTGCCACCTCAGCTCCCACTCTCTTCATATATGAGCTACTG	902
841	Db	CACCACTGGGCACTGCATGTGCCACCTCAGCTCCCACTCTCTTCATATATGAGCTACTG	900
903	Qy	GCCATGACACACTGCCATGCAATTTGGATTTCCCAACATGACACAACTGACCATGCACGC	962
901	Db	GCCATGACACACTGCCATGCAATTTGGATTTCCCAACATGACACAACTGACCATGCACGC	960
963	Qy	TCTCTCCATAAATATCTATTTAAATTTAAACATTAATTTTCATATACCTTTTTTGATGACG	1022
961	Db	TCTCTCCATAAATATCTATTTAAATTTAAACATTAATTTTCATATACCTTTTTTGATGACG	1020
1023	Qy	TGGATGCAATGGCCATCGTGTTAATAATTTGTTAAATTTGGAGTTGTAATAATAAATGAAA	1082
1021	Db	TGGATGCAATGGCCATCGTGTTAATAATTTGTTAAATTTGGAGTTGTAATAATAAATGAAA	1080
1083	Qy	GAIAAAAAAGTTGGAAAGATTTTGCATTTGTTGTGTATAAATA	1124
1081	Db	GAIAAAAAAGTTGGAAAGATTTTGCATTTGTTGTGTATAAATA	1122

RESULT 4	
AAD29066	
ID	AAD29066 standard; DNA; 1821 BP.
XX	
XX	
AC	AAD29066;
XX	
XX	
DT	07-MAY-2002 {first entry}
XX	
XX	
DE	Phaseolus vulgaris arcelin promoter.
XX	
XX	
KW	Heterologous gene expression; plant; arcelin promoter; arcelin leader;
XW	seed preferred expression cassette; ds.
XX	
XX	
OS	Phaseolus vulgaris.

XX	WO200200899-A2.
FN	
XX	03-JAN-2002.
PD	
XX	
PZ	31-MAY-2001; 2001WO-EP06298.
XX	
PR	29-JUN-2000; 2000EP-00202278.
XX	
PA	(VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX	
PI	Angenon G, De Jaeger G, Goossens A, Depicker A;
XX	WPI; 2002-139925/18.
DR	
XX	
XX	Novel seed preferred expression cassette having gene regulatory elements comprising arcelin promoter, arcelin 51 leader, and arcelin 51' end,
PT	useful for heterologous gene expression in plants.
XX	
PS	Claim 1; Page 46-47; 52pp; English.
XX	
CC	The invention relates to heterologous gene expression in plants. The invention also relates to seed preferred expression cassette having gene regulatory elements comprising arcelin promoter, arcelin 51 leader, and arcelin 51' end. This expression cassette is useful for heterologous gene expression in plants. The protein encoded by the heterologous gene is a single chain antibody variable fragment (scFv). The present sequence is Phaseolus vulgaris arcelin promoter of the invention
XX	
SQ	Sequence 1821 BP; 598 A; 308 C; 210 G; 605 T; 0 U; 0 Other;
	Query Match 91.1%; Score 1046; DB 6; Length 1821;
	Best Local Similarity 99.8%; Pred. No. 0;
	Matches 1146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 TAGATCCTTCATAGAAAATGTGTATTTCCTCATCACAGCAAAAGGGCGCAACAGTTA 60
DB	674 TAGATCCTTCAATAGAAAATGTGTATTTCCTCATCACAGCAAAAGGGCGCAACAGTTA 733
QY	61 ACAAAACAATAATTATGTTCATTTCAGATTGAAGAAGTGTAAGGAAGAAAAAGATTAAAA 120
DB	734 ACAAAACAATAATTATGTTCATTTCAGATTGAAGAAGTAAGGAAGAAAAAGATTAAAA 793
QY	121 AAAATGTCCTATCTCTTGTCTGTAATAATAATAAGAGACTTTAAACTTTTATAT 180
DB	794 AAAATGTCCTATCTCTTGTCTGTAATAATAATAAGAGACTTTAAACTTTTATAT 853
QY	181 AAATAATTGTAATTAGGTTTTCTAGTCATGAGCACCACTCAGAGCAACAGATTTCAAAGAAA 240
DB	854 AAATAATTGTAATTAGGTTTTCTAGTCATGAGCACCACTCAGAGCAACAGATTTCAAAGAAA 913
QY	241 CAATTTTGTAAACATCTTATTAGAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAACA 300
DB	914 CAATTTTGTAAACATCTTATTAGAACTTTTAGTTAAGTCTTGAAGTTAGAAATTAACA 973
QY	301 AAAAAAAAAACACGAGAAACACATAAACCCCACTACCGTCAGGTTTATCATAGGATGA 360
DB	974 AAAAAAAAAAGTACACGAGAAACACATAAACCCCACTACCGTCAGGTTTATCATAGGATGA 1033
QY	361 AATGTTTGTATATCAATTAAATATAACAACACAAAAATACATCTAAATTATAACAATAT 420
DB	1034 AATGTTTGTATATCAATTAAATATAACAACACAAAAATACATCTAAATTATAACAATAT 1093
QY	421 GTTATACATATATTTTGTAAAAAAGTTAGAGTTTTTCAAAACACTCTTAATACATGATTAG 480
DB	1094 GTTATACATATATTTTGTAAAAAAGTTAGAGTTTTTCAAAACACTCTTAATACATGATTAG 1153
QY	481 AGTTTATAGAAAATACAAATATTTAAAAATATATAATTTTAAAAAACAATCTCTAAAGTCATT 540
DB	1154 AGTTTATAGAAAATACAAATATTTAAAAATATATAATTTTAAAAAACAATCTCTAAAGTCATT 1213
QY	541 CAGATCCTCTCACACCTGTGTGATTCATTTAGTGCATGTATGTAGTACAAATCATTTGTAGTTC 600

```
Db 1214 CAGATCCTCTCACACCTGTGTGATCAATTAGTCATGTATGTAGTACAAATCATTGTAGTTC 1273
Qy 601 ACAACAGAGTAAATTAATAAATPAAGGATAAACTAGGGAATATATATAATATAATATAAT 660
Db 1274 ACAACAGAGTAAATTAATAAATPAAGGATAAACTAGGGAATATATATAATATAATATAAT 1333
Qy 661 AAAAAGGGMAATCAATTAAGATTTTGGTTCCCAATGACACACACTCACCATGCAC 720
Db 1334 AAAAAGGGMAATCAATTAAGATTTTGGTTCCCAATGACACACACTCACCATGCAC 1393
Qy 721 GCTGCCACCTCAGCTCCCTCTCTCCACACATGTCTCATGTGCACATTTTGGCTTT 780
Db 1394 GCTGCCACCTCAGCTCCCTCTCTCCACACATGTCTCATGTGCACATTTTGGCTTT 1453
Qy 781 TTCATATGACACAACTGCCCATGCAATGTTCACAGTGCAGCTCCTCTCTCTCCCATGAT 840
Db 1454 TTCATATGACACAACTGCCCATGCAATGTTCACAGTGCAGCTCCTCTCTCTCCCATGAT 1513
Qy 841 GACACACCTGGGCGATGCTGCTGCCACCTCAGCTCCCACTCTTCTCATTTATGAGCCTAC 900
Db 1514 GACACACCTGGGCGATGCTGCTGCCACCTCAGCTCCCACTCTTCTCATTTATGAGCCTAC 1573
Qy 901 TGGCAATGACACCTGCGACCTCAGCATCTCTCTCCTCTCCTTCCCATGCTGCGCAACCG 960
Db 1574 TGGCAATGACACCTGCGACCTCAGCATCTCTCTCCTCTCCTTCCCATGCTGCGCAACCG 1633
Qy 961 CTCTCTCTCCATTAATATCTATTTAAATTTAAATTAATTTAAATTAATTTAAATTAAT 1020
Db 1634 CTCTCTCTCCATTAATATCTATTTAAATTTAAATTAATTTAAATTAATTTAAATTAAT 1693
Qy 1021 COTGGATGCAATGCCATCGTCTTTTAAATTAATTTAAATTTAAATTTAAATTTAAATTAAT 1080
Db 1694 COTGGATGCAATGCCATCGTCTTTTAAATTAATTTAAATTTAAATTTAAATTTAAATTAAT 1753
Qy 1081 AAGAAAGATTGGAAAGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGG 1140
Db 1754 AAGAAAGATTGGAAAGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGG 1813
Qy 1141 TTAATGCA 1148
Db 1814 TTAATGCA 1821

RESULT 5
ABN83928
ID ABN83928 standard; DNA; 1866 BP.
XX
AC ABN83928;
XX
DT 06-SEP-2002 (first entry)
XX
DE Arcelin-4 full length promoter sequence.
XX
KW Arcelin-4; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
PN WO200250295-A2.
XX
PD 27-JUN-2002.
XX
PF 17-DEC-2001; 2001WO-US047495.
XX
DE Arcelin-4 full length promoter sequence.
XX
KW Arcelin-4; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
PN WO200250295-A2.
XX
PD 27-JUN-2002.
XX
PF 17-DEC-2001; 2001WO-US047495.
XX
PR 18-DEC-2000; 2000US-0255879P.
XX
PA (RENE-) RENESSEN LLC.
XX
PI Wang Q, Dubois P, Liang J, Oulmassov T;
XX
PI WPI; 2002-508809/54.
XX
PT New transformed or transgenic soybeans plants or cells with an Arcelin-5
```

```
PT promoter, for use as an improved dietary source of protein for humans or
PT animals, or for producing soybeans with important agricultural or
XX nutritional properties.
XX Example 1; Fig 4; 74pp; English.
XX
CC The invention relates to a transformed soybean plant cell and transgenic
CC soybean plant, both of which has a nucleic acid molecule comprising the
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
CC The transformed soybean plant cell and transgenic soybean plant are
CC useful as an improved source of dietary protein for humans and livestock.
CC These are also useful for producing soybean plants that exhibit important
CC agricultural, nutritional or pharmaceutical properties. The current
CC sequence represents an arcelin-4 full length promoter sequence
XX
XX Sequence 1866 BP; 692 A; 339 C; 209 G; 626 T; 0 U; 0 Other;
XX
Query Match 9.3%; Score 107; DB 6; Length 1866;
Best Local Similarity 100.0%; Pred. No. 1.9e-38;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 862 TGCCACCTCAGCTCCCACTCTCTCTCATTTAGAGCCTACTGGCCATGCACACTGCCACT 921
Db 1544 TGCCACCTCAGCTCCCACTCTCTCTCATTTAGAGCCTACTGGCCATGCACACTGCCACT 1603
Qy 922 CAGCACTCTCTCAGCTTCCCACTTGTCTACCTGCCAAACCGCTTCTCTC 968
Db 1604 CAGCACTCTCTCAGCTTCCCACTTGTCTACCTGCCAAACCGCTTCTCTC 1650

RESULT 6
ABN83927
ID ABN83927 standard; DNA; 1872 BP.
XX
AC ABN83927;
XX
DT 06-SEP-2002 (first entry)
XX
DE Arcelin-3 full length promoter sequence.
XX
KW Arcelin-3; promoter; plant; transgenic; soybean; agriculture; nutrition;
KW pharmaceutical; ds.
XX
OS Phaseolus vulgaris.
XX
PN WO200250295-A2.
XX
PD 27-JUN-2002.
XX
PF 17-DEC-2001; 2001WO-US047495.
XX
PR 18-DEC-2000; 2000US-0255879P.
XX
PA (RENE-) RENESSEN LLC.
XX
PI Wang Q, Dubois P, Liang J, Oulmassov T;
XX
PI WPI; 2002-508809/54.
XX
PT New transformed or transgenic soybeans plants or cells with an Arcelin-5
PT promoter, for use as an improved dietary source of protein for humans or
PT animals, or for producing soybeans with important agricultural or
XX nutritional properties.
XX Example 1; Fig 4; 74pp; English.
XX
CC The invention relates to a transformed soybean plant cell and transgenic
CC soybean plant, both of which has a nucleic acid molecule comprising the
CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
CC The transformed soybean plant cell and transgenic soybean plant are
CC useful as an improved source of dietary protein for humans and livestock.
CC These are also useful for producing soybean plants that exhibit important
CC agricultural, nutritional or pharmaceutical properties. The current
XX sequence represents an arcelin-4 full length promoter sequence
```


XX (RENE-) RENESSEN LLC.
 XX Wang Q, Dubois P, Liang J, Oulmassov T;
 PI WPI; 2002-508809/54.
 XX
 XX New transformed or transgenic soybeans plants or cells with an Arcelin-5
 PT promoter, for use as an improved dietary source of protein for humans or
 PT animals, or for producing soybeans with important agricultural or
 PT nutritional properties.
 XX
 XX Example 1; Fig 1; 74pp; English.
 XX
 XX The invention relates to a transformed soybean plant cell and transgenic
 CC soybean plant, both of which has a nucleic acid molecule comprising the
 CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
 CC The transformed soybean plant cell and transgenic soybean plant are
 CC useful as an improved source of dietary protein for humans and livestock.
 CC These are also useful for producing soybean plants that exhibit important
 CC agricultural, nutritional or pharmaceutical properties. The current
 CC sequence represents an arcelin-4 promoter sequence fragment
 XX
 XX Sequence 316 BP; 113 A; 37 C; 37 G; 129 T; 0 U; 0 Other;
 SQ
 Query Match 2.3%; Score 26; DB 6; Length 316;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 662 AAAAGGGGAAATCAATAGATT 687
 DB 197 AAAAGGGGAAATCAATAGATT 222
 RESULT 10
 ID ABN83924 standard; DNA; 322 BP.
 AC ABN83924;
 XX
 XX 06-SEP-2002 (first entry)
 DT
 DE Arcelin-3 promoter sequence fragment.
 XX
 XX Arcelin-3; promoter; plant; transgenic; soybean; agriculture; nutrition;
 KW pharmaceutical; ds.
 XX
 XX Phaseolus vulgaris.
 OS
 XX WO200250295-A2.
 PN
 XX 27-JUN-2002.
 PD
 XX 17-DEC-2001; 2001WO-US047495.
 PF
 XX 18-DEC-2000; 2000US-0255879P.
 PR
 XX (RENE-) RENESSEN LLC.
 XX
 XX Wang Q, Dubois P, Liang J, Oulmassov T;
 PI WPI; 2002-508809/54.
 XX
 XX New transformed or transgenic soybeans plants or cells with an Arcelin-5
 PT promoter, for use as an improved dietary source of protein for humans or
 PT animals, or for producing soybeans with important agricultural or
 PT nutritional properties.
 XX
 XX Example 1; Fig 1; 74pp; English.
 PS
 XX The invention relates to a transformed soybean plant cell and transgenic
 CC soybean plant, both of which has a nucleic acid molecule comprising the
 CC Phaseolus vulgaris exotic genotype G02771 Arcelin-5 promoter sequence.
 CC

CC The transformed soybean plant cell and transgenic soybean plant are
 CC useful as an improved source of dietary protein for humans and livestock.
 CC These are also useful for producing soybean plants that exhibit important
 CC agricultural, nutritional or pharmaceutical properties. The current
 CC sequence represents an arcelin-3 promoter sequence fragment
 XX
 XX Sequence 322 BP; 114 A; 37 C; 38 G; 133 T; 0 U; 0 Other;
 SQ
 Query Match 2.3%; Score 26; DB 6; Length 322;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 662 AAAAGGGGAAATCAATAGATT 687
 DB 197 AAAAGGGGAAATCAATAGATT 222
 RESULT 11
 ID AAN90025 standard; DNA; 1542 BP.
 XX AAN90025;
 XX
 XX 25-MAR-2003 (revised)
 DT 17-DEC-2001 (revised)
 DT 01-NOV-1989 (first entry)
 DE DNA encoding antigenic circumsporozoite protein of Plasmodium malariae.
 XX
 XX Antigenic; vaccine; Plasmodium malariae; circumsporozoite protein.
 KW
 XX Plasmodium malariae.
 OS
 XX
 XX Key Location/Qualifiers
 FH CDS 98..1383
 FT /*tag= a
 FT misc_feature 424..1036
 FT /*tag= b
 XX
 XX USN7238746-N.
 PN
 XX 21-MAR-1989.
 PD
 XX 21-MAR-1989; 89US-00238746.
 PF
 XX 31-AUG-1988; 88US-00238746.
 PR
 XX (USSH) NAT INST OF HEALTH.
 XX
 XX WPI; 1989-172958/23.
 DR
 XX
 XX Antigenic protein for vaccination against malaria - encoded by cloned
 PT gene coding for plasmodium malariae circumsporozoite protein.
 PT
 XX Disclosure; Fig 1; 11pp; English.
 PS
 XX
 XX DNA encoding antigenic circumsporozoite protein of Plasmodium malariae.
 CC Used as a vaccine against malaria. Nucleotides 424 - 1036 make up the
 CC immunodominant region. (Note: Revised entry submitted to correct the
 CC patent number format of US Government-owned NTIS applications to prevent
 CC clashes with ongoing US granted patent numbers. For further information
 CC please visit the Derwent web site at
 CC www.derwent.com/dwpi/updates/ntis.us.html.) (Updated on 25-MAR-2003 to
 CC correct PD field.) (Updated on 25-MAR-2003 to correct PP field.) (Updated
 CC on 25-MAR-2003 to correct PR field.)
 XX
 XX Sequence 1542 BP; 618 A; 228 C; 344 G; 352 T; 0 U; 0 Other;
 SQ
 Query Match 2.2%; Score 25; DB 1; Length 1542;
 Best Local Similarity 100.0%; Pred. No. 0.39;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 634 GAATATATATATATACAAATTAA 658

Db 1513 GAATATATATATATATACAAATTA 1489
|||||
RESULT 12
ABZ80122
ID ABZ80122 standard; DNA; 24 BP.
XX
AC ABZ80122;
XX
DT 22-MAY-2003 (first entry)
XX
DE Arcelin 5 promoter PCR primer SEQ ID NO:79.
XX
KW Gamma-tocopherol methyltransferase; methyltransferase; tocopherol;
KW alpha-tocopherol; alpha-tocotrienol; stress; oxidative stress tolerance;
KW oxygen; ozone; UV tolerance; ultra violet tolerance; cold tolerance;
KW fungal pathogen; microbial pathogen; plant; enzyme; PCR primer; ss.
XX
OS Synthetic.
XX
FN WO2003016482-A2.
XX
PD 27-FEB-2003.
XX
PF 16-AUG-2002; 2002WO-US026047.
XX
PR 17-AUG-2001; 2001US-0312758P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Van Eenennaam A, Valentin HE, Karunanandaa B, Hao M, Aasen E;
PI Levering C;
XX
DR WPI; 2003-268314/26.
XX
PT Novel substantially purified tocopherol protein and nucleic acid encoding
PT the same, useful for producing a plant having a seed with increased alpha
PT -tocopherol level.
XX
PS Example 6; Page 94; 218pp; English.
XX
CC The present invention describes a substantially purified tocopherol
CC protein (I). Also described is a method (M) for reducing expression of
CC methyltransferase 1 (MT1) or gamma-tocopherol methyltransferase (GMT) in
CC a plant, by transforming a plant with a nucleic acid molecule having an
CC exogenous promoter region which functions in plant cells to cause the
CC production of an mRNA molecule, and growing the transformed plant. (M) is
CC useful for increasing the gamma-tocopherol content. A polynucleotide (II)
CC encoding (i) can be used for producing a plant having a seed with an
CC increased alpha-tocopherol or alpha-tocotrienol level. (II) is useful for
CC modulating the reduction of the expression, expression, overexpression of
CC (i) in a transformed plant to provide tolerance to a variety of stresses,
CC e.g. oxidative stress tolerance such as to oxygen or ozone, UV tolerance,
CC cold tolerance, or fungal/microbial pathogen tolerance. (II) is useful
CC for obtaining other nucleic acid molecules or homologues from the same
CC species, to screen cDNA or genomic libraries, to isolated promoters of
CC cell enhanced, cell specific, tissue enhanced, tissue specific,
CC developmentally or environmentally regulated expression profiles, as
CC markers, for detecting single nucleotide polymorphisms, and to determine
CC the level of (i) in a plant or pattern of expression of (i) encoded in
CC part or whole by (II). The present sequence represents a PCR primer for
CC an arcelin 5 promoter sequence, which is used in an example from the
CC present invention
XX
SQ Sequence 24 BP; 2 A; 12 C; 3 G; 7 T; 0 U; 0 Other;
Query Match 2.1%; Score 24; DB 7; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 812 CCACGTGAGCTCTTCTCTTCCC 835
|||||

Db 1 CCACGTGAGCTCTTCTCTTCCC 24
RESULT 13
AAS62538/c
ID AAS62538 standard; CDNA; 1224 BP.
XX
AC AAS62538;
XX
DT 14-FEB-2002 (first entry)
XX
DE cDNA sequence #325 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antirheumatic; ss.
XX
OS Homo sapiens.
XX
FN WO200177291-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US010485.
XX
PR 06-APR-2000; 2000US-0195604P.
XX
PA (GENY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-010900/01.
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease.
XX
PS Claim 1; Page 246; 391pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides a
CC method for producing proteins from these polynucleotide sequences. The
CC proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are useful
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),
CC immune deficiency disorders (e.g. severe combined immunodeficiency
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders
CC (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of
CC the invention are also useful in gene therapy. AAS62214-AAS62838
CC represent the cDNA sequences of the invention that encode for novel human
CC secreted proteins
XX
SQ Sequence 1224 BP; 395 A; 205 C; 204 G; 420 T; 0 U; 0 Other;
Query Match 2.1%; Score 24; DB 6; Length 1224;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 501 TTTAAATAATATATTTTAAAAA 524
|||||
Db 763 TTTAAATAATATATTTTAAAAA 740
Query Match 2.1%; Score 24; DB 6; Length 1224;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 501 TTTAAATAATATATTTTAAAAA 524
|||||
Db 763 TTTAAATAATATATTTTAAAAA 740
RESULT 14
ABA46475/c
ID ABA46475 standard; DNA; 400 BP.
XX
AC ABA46475;
XX

```
XX 01-FEB-2002 (first entry)
XX Human breast cell single exon nucleic acid probe #5170.
XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer; ss.
XX Homo sapiens.
XX WO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000662.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.
XX Claim 1; SEQ ID NO 5170; 327pp + Sequence Listing; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX prognosing breast disease. Gene expression analysis is useful for
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a single exon
XX nucleic acid probe of the invention. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 400 BP; 89 A; 85 C; 94 G; 132 T; 0 U; 0 Other;
XX
XX Query Match 2.0%; Score 23; DB 4; Length 400;
XX Best Local Similarity 100.0%; Pred. No. 3.2;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 490 AAATACAAATATTTAAAAAATAT 512
XX Db 250 AAATACAAATATTTAAAAAATAT 228
XX
XX RESULT 15
XX AAK30701/c
XX ID AAK30701 standard; DNA; 400 BP.
XX AC AAK30701;
XX XX 06-NOV-2001 (first entry)
```

```
XX Human bone marrow expressed single exon probe SEQ ID NO: 5258.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-489900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 5258; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention
XX Sequence 400 BP; 89 A; 85 C; 94 G; 132 T; 0 U; 0 Other;
XX
XX Query Match 2.0%; Score 23; DB 4; Length 400;
XX Best Local Similarity 100.0%; Pred. No. 3.2;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 490 AAATACAAATATTTAAAAAATAT 512
XX Db 250 AAATACAAATATTTAAAAAATAT 228
XX
XX Search completed: June 2, 2004, 04:59:46
XX Job time : 547 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2004, 02:15:29 ; Search time 4648 Seconds
(without alignments)
10705.207 Million cell updates/sec

Title: US-10-015-637-1
Perfect score: 1148
Sequence: 1 taggatacttcaatagaaaa.....agagagtgtgtaaatgca 1148

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_higo_hum.*
40: em_higo_mus.*
41: em_higo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	1148	6	AX463281 Sequence
2	1148	100.0	1832	6	AX463294 Sequence
3	1144.8	99.7	1821	6	AX343909 Sequence
4	1144.8	99.7	3900	6	PVARG5X1
5	1122	97.7	1122	6	AX463282 Sequence
6	796.6	68.4	1866	6	AX463293 Sequence
7	795	69.3	1872	6	AX463292 Sequence
8	274.6	23.9	4564	8	PHVARCIA
9	139.4	12.1	2288	8	AF193029
10	118.2	10.3	1768	8	PVPDLEC1
11	109.2	9.5	22243	3	PFVAR23A
12	108.4	9.4	1395	8	PVDLEC1
13	108.4	9.4	1441	8	PHVDLECA
14	95.8	8.3	24611	2	AC111404
15	95.2	8.3	93791	2	AC138073
16	95	8.3	125933	2	AC146394
17	92	8.0	1689	8	PHVLECT
18	92	8.0	4846	8	AF325188
19	91.8	8.0	111861	9	AC069435
20	91.8	8.0	170827	2	AC125567
21	91.6	8.0	1992	8	PVPDLEC2
22	91.2	7.9	175544	2	AC117342
23	90.6	7.9	258658	3	AE014832
24	90	7.8	8056	6	AX599046
25	90	7.8	110000	2	PFMAL8P1_12
26	89.8	7.8	810	8	AF592180
27	89.8	7.8	14867	3	AE001398
28	89.8	7.8	136888	9	AL353783
29	89.6	7.8	103344	9	HS1100E15
30	89.6	7.8	141275	2	BX510640
31	89.6	7.8	171333	5	BX248097
32	89.6	7.8	249867	2	AC127704
33	88.8	7.7	108908	3	PFMAL3P8
34	88.8	7.7	195620	2	BX088600
35	88.6	7.7	250029	3	AE014820
36	88.6	7.7	250743	3	AE014836
37	87.4	7.6	1434	8	ADJ49822
38	87.4	7.6	170880	2	BX640469
39	87.4	7.6	250029	3	AE014830
40	87.4	7.6	1453	8	AJ591978
41	87.2	7.6	164640	2	BX005461
42	87.2	7.6	168558	2	BX322549
43	87.2	7.6	168558	2	BX322549
44	87	7.6	3683	6	AX598999
45	87	7.6	136657	9	AC068600

ALIGNMENTS

RESULT 1
AX463281
LOCUS AX463281 1148 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 1 from Patent WO0250295.
ACCESSION AX463281
VERSION AX463281.1 GI:21886232
KEYWORDS
SOURCE Phaseolus vulgaris
ORGANISM Phaseolus vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
REFERENCE 1
AUTHORS Oulmassov,T., Wang,Q., Dubois,P. and Liang,J.

Pred. No. is the number of results predicted by chance to have a

[Z] / CITATION=

Query Match	Best Local Similarity	Matches 11a6;
1	TAC	1
674	TAC	674
61	AC	61
734	AC	734
121	AA	121
794	AA	794
181	AA	181

99.7%;	Score 1144.8;	DB 8;	Length 3900;
ty 99.8%;	Prod. No. 2.2e-172;		
ervative	0; Mismatches 2;	Indels 0;	Gaps 0;
CCTTCAATAGAAAATGTGTTATTTCTTCATCACAGACAAAGGGCGAACAGTTA	60		
CTTCAATAGAAAATGTGTTATTTCTTCATCACAGACAAAGGGCGAACAGTTA	733		
AAAAATTATGTTTCATTTGAGATTAAAGGAAGTAAAGGAAGAAAAAGATTAAAA	120		
AAAAATTATGTTTCATTTGAGATTAAAGGAAGTAAAGGAAGAAAAAGATTAAAA	793		
CCCTATCTCTTTGTTCTGTAATAATATATAGAGACTTAAACCTTTTAAATAT	180		
CCCTATCTCTTTGTTCTGTAATAATAATATAGAGACTTAAACCTTTTAAATAT	853		
GTAAATTAGGTTTTCATGCTATGACCACTCAGAGACAAAGATTTCAGAGAAA	240		

Web

[illegible]

RESULT 5				
AX463282				
LOCUS	AX463282	1122 bp	DNA	linear
DEFINITION	Sequence 2 from Patent WO0250295.			
ACCESSION	AX463282			
				PAT 15-JUL-2002

VERSION	AX463282.1	GI:21886233
KEYWORDS	Phaseolus vulgaris	
SOURCE	Phaseolus vulgaris	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; fabales; fabaceae; Papilionoideae; Phaseoleae; Phaseolus.	
REFERENCE	1	
AUTHORS	Olummasov, T., Wang, Q., Dubois, P. and Liang, J.	
TITLE	Arcein-5 promoter and 27 uses thereof	
JOURNAL	Patent: WO 0250295-A 2 27-JUN-2002;	
FEATURES	RENNSEN LLC (US)	
source	Location/Qualifiers	
ORIGIN	1..1122	
	/organism="Phaseolus vulgaris"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:3885"	
Query Match	97.7%;	Score 1122; DB 6; Length 1122;
Best Local Similarity	100.0%;	Pred. No. 1.le-168;
Matches 1122;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	3	GGATCCTTCAATAGAAATGTGTTATTTCTCATCCAGACAAAGGGGCAACAGTTAAAC 62
Db	1	GGATCCTTCAATAGAAATGTGTTATTTCTCATCCAGACAAAGGGGCAACAGTTAAAC 60
QY	63	AAAAACAATTTATGTTTCATTTTCAGATTAAAGAGSTAGAGAGAAAGATTAATAAAA 122
Db	61	AAAAACAATTTATGTTTCATTTTCAGATTAAAGAGSTAGAGAGAAAGATTAATAAAA 120
QY	123	AATGCTCTATCTCTTGTGTTTCGTTAATTAATTAATAGAGACTTAAACTTTTAAATAA 182
Db	121	AATGCTCTATCTCTTGTGTTTCGTTAATTAATTAATAGAGACTTAAACTTTTAAATAA 180
QY	183	TAAATTGTAATTAGTGTTCCTAGTCATGACACCACTCAGAGACAAAGATTTCACAGAAACA 242
Db	181	TAAATTGTAATTAGTGTTCCTAGTCATGACACCACTCAGAGACAAAGATTTCACAGAAACA 240
QY	243	ATTTTGTGTAACACTCTATTAGAGAACTTTTAGTTAGTCTCTGAGTTAGATTAAACAAA 302
Db	241	ATTTTGTGTAACACTCTATTAGAGAACTTTTAGTTAGTCTCTGAGTTAGATTAAACAAA 300
QY	303	AAAAATTACACACGAGAAACACAATTAACCCCACTACCGTCAGGTTATCATAGAGATGAAA 362
Db	301	AAAAATTACACACGAGAAACACAATTAACCCCACTACCGTCAGGTTATCATAGAGATGAAA 360
QY	363	TGTTTGTATATCATTTAAATAATAACACACAAAAATACATCTAATTATTAACAATATATGT 422
Db	361	TGTTTGTATATCATTTAAATAATAACACACAAAAATACATCTAATTATTAACAATATATGT 420
QY	423	TATACATATATTTTGTGTAAGAACTTTAGAGTTTTTCAGAACATTTCTTAATACATGATTAGAG 482
Db	421	TATACATATATTTTGTGTAAGAACTTTAGAGTTTTTCAGAACATTTCTTAATACATGATTAGAG 480
QY	483	TTTATAGAAATACAAATATTTTAAAAAATAATATTTTAAAAAAACAATCTTAAAGTCATTCA 542
Db	481	TTTATAGAAATACAAATATTTTAAAAAATAATATTTTAAAAAAACAATCTTAAAGTCATTCA 540
QY	543	GATCCTCTCACACCTGTGTCATCAATTTAGTCATGTATGTAGTACAAATCATTTGTAGTTCCAC 602
Db	541	GATCCTCTCACACCTGTGTCATCAATTTAGTCATGTATGTAGTACAAATCATTTGTAGTTCCAC 600
QY	603	AACAGCTTAAATAATAATAGGTAATAACTAGGGATATATATATATATATACAAATTAATAA 662
Db	601	AACAGGTAATAATAATAATAGGTAATAACTAGGGATATATATATATATATACAAATTAATAA 660
QY	663	AAAGGGGAAATCAAAATTAGAAATTTTGATTTCCCAACATGACAACTCACCATGACGCG 722
Db	661	AAAGGGGAAATCAAAATTAGAAATTTTTGATTTCCCAACATGACAACTCACCATGACGCG 720
QY	723	TGCACCTTCAGCTCCCTCCCTCTCCACACATGTCATGTCTACITTTTCGACTTTGGCTTTT 782